

Washington University in St. Louis Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Winter 12-15-2018

The Effect of Talker and Contextual Variability on Memory for Words in Sentences

Nichole Runge

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds



Part of the [Cognitive Psychology Commons](#), and the [Other Psychology Commons](#)

Recommended Citation

Runge, Nichole, "The Effect of Talker and Contextual Variability on Memory for Words in Sentences" (2018). *Arts & Sciences Electronic Theses and Dissertations*. 1726.

https://openscholarship.wustl.edu/art_sci_etds/1726

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological and Brain Sciences

Dissertation Examination Committee:
Mitchell Sommers, Chair
Joe Barcroft
Mark McDaniel
Henry Roediger
Kristin Van Engen

The Effect of Talker and Contextual Variability on Memory for Words in Sentences
by
Nichole Runge

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2018
St. Louis, Missouri

© 2018, Nichole Runge

Table of Contents

List of Figures	iv
List of Tables	v
Acknowledgments	vi
Abstract of the Dissertation	viii
Chapter 1: Introduction	1
Talker Variability	3
Memory for Known Words	3
Memory for Novel Words	8
Talker Variability in Experiment 1	13
Contextual Variability	14
Unrelated Contexts	14
Related Contexts	17
Less-Related Contexts	19
Overview of Contextual Variability Findings	23
Contextual Variability in Experiment 1	24
Chapter 2: Experiment 1	27
Method	29
Participants	30
Materials	31
Procedure	35
Results	36
Target Word Recall	36
Intrusions	39
Posttest Questionnaire Responses	40
Discussion	41
Contextual Variability Findings from Experiment 1	42
Chapter 3: Experiment 2	45
Change in the Presentation Modality	45
Interaction of Presentation Modality and Contextual Variability	45
Main Effect of Presentation Modality	46

Addition of Working Memory Measures	47
Study-Phase Retrieval Theory	48
Difficulty from Increased Lag	49
Difficulty from Intervening Task Demands	49
Difficulty from Contextual Variability (Present Experiment).....	52
Method	53
Participants	53
Materials & Procedure	54
Results	55
Target Word Recall (Experiment 2 only)	56
Questionnaire Responses (Experiment 2 only)	59
Target Word Recall (Experiments 1 & 2)	59
Intrusions (Experiments 1 & 2)	62
Discussion	62
Chapter 4: General Discussion	65
Summary of Findings	65
Talker Variability	66
Contextual Variability	68
Presentation Modality	69
Test Expectations.....	71
Spaced Versus Massed Stimuli Presentation.....	72
General Conclusions	74
References.....	76
Appendices.....	82

List of Figures

Figure 1. Model of the effect of variable acoustic input (Barcroft & Sommers, 2005).....	9
Figure 2. Cued definition-to-word recall (Runge et al., 2017; L1 talker variability experiment). 11	
Figure 3. Cued definition-to-L2 recall (Runge et al., 2017; written definition Spanish experiment).	12
Figure 4. Schematic of study phase procedure (Experiment 1).	33
Figure 5. Effects of number of talkers and context sentences on the proportion of target words recalled (Experiment 1).....	37
Figure 6. Predicted episodic memory performance as a function of working memory ability and task difficulty (Bui et al., 2013).	51
Figure 7. Scatterplot of working memory capacity and target word recall in the constant- and variable-context conditions (Experiment 2).....	58
Figure 8. Effects of presentation modality and number of context sentences on the proportion of target words recalled (Experiments 1 & 2).	61

List of Tables

Table 1. Overview of contextual variability experiments.....	23
Table 2. Example stimuli in the four study conditions (Experiment 1).....	34
Table 3. Means for the proportion of target words recalled based on the amount of talker and contextual variability during study (Experiment 1).	37
Table 4. Means for the proportion of target words recalled in constant-talker conditions based on number of context sentences and presentation modality (Experiments 1 & 2).	60

Acknowledgments

First, I would like to thank my adviser and mentor Mitch Sommers for his guidance and support throughout graduate school. I would also like to thank my other core dissertation committee members, Roddy Roediger and Mark McDaniel, for being on this committee, as well as my masters and subject matter examination committees. Thanks also go out to my other committee members, Kristin Van Engen and Joe Barcroft, who have been valued collaborators and mentors throughout my time at WashU. In addition, I would like to thank the person who first introduced me to psychology many years ago—Debby Burke.

Thanks go out to the research assistants that helped me collect data for this dissertation: Ariadne Bazigos, Kyra Reu, and Adam Isaacs. I would also like to acknowledge the other members of the Speech & Hearing Lab for being great colleagues with helpful feedback.

To my parents, Laurie and Frank, thank you for instilling the importance of education at a young age. I would also like to thank Scott Johnson for his recommendations regarding education, and I want to give a shout out to all my siblings—Vada, Frankie, and Peer. Finally, I want to say thank you to my husband Jeremy for his constant support.

Nichole Runge

Washington University in St. Louis

December 2018

Dedicated to my Grandma, ‘Nannie’ Ruby Runge.

Abstract of the Dissertation
The Effect of Talker and Contextual Variability on Memory for Words in Sentences
by
Nichole Runge
Doctor of Philosophy in Psychological & Brain Sciences
Washington University in St. Louis, 2018
Professor Mitchell Sommers, Chair

Previous research has found that adding different forms of variability during study can affect later memory at test. For example, having words spoken by different talkers has been shown to improve recall of known and novel words (Goldinger et al., 1999; Barcroft & Sommers, 2005), and varying the cues in cue-target related word pairs has been found to improve recall of the targets (Glenberg, 1979; Bevan et al., 1966). It was unclear, however, whether benefits of variability would extend to more naturalistic stimuli, such as sentences, which have higher working memory demands. The present set of experiments investigated how talker and contextual variability, both individually and combined, affect free recall of target words that appear in semantically-related sentences.

Target words were sentence-final items, and all stimuli in Experiment 1 were presented auditorily and orthographically. For each participant, targets appeared in one of the following four conditions: the same sentence spoken three times by the same person (no variability), three different sentences spoken by the same person every time (contextual variability), the same sentence spoken by three different talkers once each (talker variability), or a different sentence spoken by a different talker at each of three exposures (combined contextual and talker variability). Conditions with contextual variability resulted in significantly worse memory performance than constant-context conditions. There was no significant effect of talker variability and no significant interaction between talker and contextual variability.

Experiment 2 further investigated the unexpected negative effect of contextual variability observed in Experiment 1 by changing the presentation modality to auditory-only (all with a constant talker). The switch from combined auditory-orthographic to auditory-only presentations was designed to both decrease working memory demands and encourage processing of the sentence as it unfolded in time. In addition, working memory measures were collected in order to test two predictions—that working memory would be a significant predictor of target word recall and that it would be a significantly better predictor in the variable-context compared to the constant-context condition. No recall differences between the constant- and variable-context conditions were found, but there was a significant positive relationship of working memory on target word recall. Lastly, although positive relationship of working memory on target word recall was stronger in the variable- than the constant-context condition, the interaction was not statistically significant.

These findings suggest that the benefits of talker and contextual variability that have previously been found for lists of words or word pairs (e.g., Glenberg, 1979; Barcroft & Sommers, 2005) do not necessarily extend to semantically-related sentences. The results are discussed with regard to working memory demands and how this may interact with variability.

Chapter 1: Introduction

Memory is a fundamental aspect of everyday life. For example, people must remember the names of others, what to buy at the grocery store, and lessons from class. A number of factors have been identified that can affect one's memory, including the way in which the information is encoded (e.g., Bower, 1970), the length of time it is stored (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993), and the similarity between encoding and retrieval (e.g., Godden & Baddeley, 1975).

With respect to the way in which information is encoded, there are several manipulations that can influence memory. For example, instructions regarding a participant's interaction with stimuli during study can alter later memory for that information. When people are told to visually imagine word pairs interacting, memory performance is better than if they had merely rehearsed the word pairs (e.g. Bower, 1970). Similarly, if someone is asked to decide if a word fits in a particular category (e.g. Is it an animal?), their memory performance is typically better than if they had been asked to judge if the word was written in all capital letters (e.g. Craik & Tulving, 1975).

In addition to explicit encoding instructions and study phase requirements, the way in which stimuli are presented can also impact later memory of those items. When participants are shown part of a related word pair (rapid-f___) and must generate the target word (fast), recall of the target is better than if they had simply been shown the intact word pair (e.g., Slamecka & Graf, 1978). The spacing of stimuli during study also influences memory such that stimuli studied during a single long session are typically remembered more poorly than stimuli that are studied in multiple shorter sessions (e.g., Bloom & Shuell, 1981).

In addition to the manipulations mentioned above, the extent of variability in the stimuli can also influence memory. Two types of stimulus variability that have been examined in terms

of their effects on later memory performance are talker and contextual variability. The first type of variability—talker variability—occurs when words are spoken by different talkers compared to the same talker. Stimuli can also be varied by changing the context in which they appear. Context can refer to anything separate from, but related to, the stimulus itself, but in the present dissertation it is used more narrowly to refer to the word or words surrounding a target item.

I will begin this dissertation by reviewing previous work regarding the effect of talker and contextual variability on memory. Some of these studies have found a benefit of variability (e.g., Bevan et al., 1966; Goldinger et al., 1991; Barcroft & Sommers, 2005), which is hypothesized to result from increased variability creating a broader lexical representation and/or additional retrieval routes. However, the majority of this research has been carried out using lists of words or word pairs, and there is little work examining the effects of talker or contextual variability on memory for words that appear in sentences. Given that human communication typically occurs at the sentence level or higher, understanding the effect of variability on memory for items that appear in a sentence is an important next step in this line of research. Positive findings also have the potential to be applied to real-life situations. For example, if talker variability in the context of a sentence improves memory for a target item, then health-care professionals (such as the doctor, nurse, and pharmacist) can all utter a keyword to improve memory for that item compared to the same professional repeating the keyword in sentences.

From a theoretical perspective, identifying and combining words in sentence-length material requires greater processing resources than identifying isolated words, and this increase in processing demands could fundamentally alter how variability affects memory performance. For example, processing of the sentence as a whole may compete for resources that would otherwise support sustained rehearsal of the target. On the other hand, sentential processing, especially in one's first language, is a common occurrence in everyday life and may not be

cognitively costly enough to negatively affect the processing of variable stimuli. In the present dissertation, I aim to better understand this line of inquiry by investigating the effect of talker and contextual variability on memory for target words that appear in high predictability sentences. Specifically, I compare the number of target words recalled when they appear in repeated or varied sentence contexts and when they are spoken by the same or different talkers. The results of these experiments provide information about boundary conditions regarding the benefits of variable input and how working memory demands may interact with variable input to affect memory.

Talker Variability

Speech contains both lexical and indexical information. Lexical information includes the acoustic and linguistic information used to identify the utterance, whereas indexical information conveys talker characteristics, such as gender, approximate age, health, dialect, and emotional state (Pisoni, 1997). When a person says “I went to the park yesterday,” a listener gains semantic information about the speaker’s whereabouts the day before, as well as some information about the speaker such as those mentioned above. Introducing variability in the form of different talkers allows researchers to keep the lexical content the same across conditions and examine effects of variability in indexical properties.

Memory for Known Words

Most of the initial work on variability of indexical properties was conducted to examine the effects of talker variability on first-language (L1) speech perception. In these studies, identification of words spoken by multiple talkers was compared to identification of those same

words spoken by a single talker. For example, Mullennix, Pisoni, and Martin (1989) exposed participants to the carrier phrase “Say the word _____ for me” in noise, with different words presented in the blank. Two types of lists were presented in a blocked format. In one list type, the same talker spoke all 15 of the carrier phrases and words, while in the other list type, the talker producing the carrier phrase and word was rotated through 15 different speakers (7 male and 8 female). After each trial, participants named the target word as quickly and accurately as possible into a microphone that recorded response latencies. Results indicated that participants were faster and more accurate at naming the words from lists spoken by a single talker compared with multiple talkers, suggesting there is a perceptual cost to identifying words with variable input in the form of different talkers. Similar findings have also been reported for children (Ryalls & Pisoni, 1997), younger adults (Sommers, Nygaard, & Pisoni, 1994; Ryalls & Pisoni, 1997), and older adults (Sommers, 1997; Kirk, Pisoni, & Miyamoto, 1997).

At the time, the predominant explanation for the benefit of same-talker speech was that variability in indexical features was a source of noise that had to be discarded before the listener could match the acoustic signal they heard to the prototypical representation stored in long-term memory (Pisoni, 1997; Mullinnex et al., 1989). Although no two utterances, even by the same person, will ever be acoustically identical, there is more overlap and consistency in same-talker speech compared to multiple-talker speech. Thus, the reason that identification was believed to be easier with same-talker speech was that encoding demands were reduced because listeners did not have to continually adjust to the acoustic characteristics of new talkers.

Another possible explanation for multiple-talker speech being more cognitively costly to perceive than same-talker speech is that listeners are actually encoding, rather than discarding, the indexical information along with the lexical information. One way to determine whether the poorer performance with multiple-talker lists was due to discarding or encoding of indexical

information is through memory studies. If indexical features need to first be discarded before identifying the word, then the increased perceptual processing of multiple talker-speech would be expected to impair memory. This would be due to the fact that the process of actively discarding indexical information would compete for resources that would otherwise be used for rehearsal and other processes that support memory. On the other hand, if multiple-talker speech is perceived more slowly due to the indexical features being encoded and stored in memory, then one might expect that multiple-talker speech would produce better memory than same-talker speech if listeners are given sufficient time to store the talker information. The rationale for this prediction is, in the case of a single talker, only the lexical information distinguishes one word from another. In contrast, both the lexical and indexical information differ between words in the case of multiple-talker speech, so variable indexical information may provide an additional source of information to aid recall.

One of the first studies to investigate the effect of talker variability on memory predicted that multiple talkers would improve memory for words presented in the recency portion of word lists due to distinct echoic memory traces associated with the different talkers (Watkins & Watkins, 1980). Participants were exposed to recordings of lists containing 16 two-syllable words at a rate of 1000 ms per word (Watkins & Watkins, 1980). For half of the lists, all of the words were spoken by the same person, whereas for the other half of the lists, the words were alternately spoken by one male and one female. After each list was presented, participants were asked to write down all of the words from the previous list that they remembered in any order. Memory for the words did not differ significantly between the one- and two-talker lists. Although Watkins and Watkins (1980) did not find better recall for the two-talker lists as they had predicted, one potential explanation is that the presentation rate was too fast for the participants to encode the indexical information, which could result in equivalent recall for the

one- and two-talker lists. On average, durations for two-syllable words are approximately 1000 ms, so it is likely that the two-syllable words used in the experiment were presented with little or no pause between trials. Therefore, participants only had a small window of time to encode the indexical information. Thus, it is possible that at a presentation rate that allows sufficient time for processing and encoding of the indexical information, a talker variability benefit to memory would be found.

In a later study by Martin, Mullennix, Pisoni, and Summers (1989), the presentation rate was increased, but recall was also changed from free to serial. This switch in recall type requires additional processing resources as not only the word, but also the word's position, must be explicitly encoded. During study, participants heard a list of ten monosyllabic words spoken either by the same talker or by ten different talkers. There was an interword interval of 1500 ms for all of the lists, and immediately following the presentation of each list, there was a 60-second period during which participants were asked to serially recall the ten words from the list. The authors predicted that participants would correctly recall more of the words from the same-talker lists compared to the multiple-talker lists because multiple-talker lists are more perceptually difficult and require more processing. If there is an increase in perceptual processing, in this case from the varied indexical information in the multiple-talker list, they hypothesized that this would result in less time and fewer resources for processing, rehearsing, and encoding the lexical information of the item. Results supported this prediction, and they found that more words in the primacy portion of same-talker lists were correctly recalled compared to multiple-talker lists.

In a follow-up experiment, however, Goldinger et al. (1991) suggested that participants in the Martin et al. (1989) study did not have sufficient time to encode the indexical talker information and, therefore, were unable to benefit from the multiple talkers. Specifically, they proposed that given sufficient time to bind the lexical and indexical information, participants

could benefit from multiple talkers because the varied indexical information could provide an additional retrieval cue. To test this hypothesis, Goldinger et al. (1991) replicated the Martin et al. (1989) study but manipulated the interword interval (IWI) between list items (range: 250 ms - 4000 ms). Goldinger et al. (1991) suggested that listeners would be unable to bind the lexical and indexical information at the shorter IWIs, and therefore, same-talker lists would produce better recall performance than multiple-talker lists. In contrast, for the longer IWIs, listeners were predicted to have sufficient time to encode and bind indexical and lexical information together, and the varied indexical information was expected to serve as an additional retrieval cue. Therefore, Goldinger et al. (1991) hypothesized better performance for the multiple-talker lists than for the same-talker lists in trials with longer IWIs. Consistent with this hypothesis, they found that for the shorter IWIs of 250 ms and 500 ms, results replicated earlier findings (Martin et al., 1989) such that words from the primacy portion of same-talker lists were remembered more accurately than those from multiple-talker lists. However, they also confirmed that at the longer IWIs of 2000 ms and 4000 ms, the findings reversed and that words in the primacy portion of the multiple-talker lists were remembered more accurately than those in the same-talker lists. These findings from Goldinger et al (1991) suggest that with enough processing time, indexical information from multiple talkers may bind with the lexical information to create a broader, more accessible lexical representation, leading to improved memory performance. Although there is a cognitive cost to identifying multiple- compared with same-talker speech, it appears that this cost can be outweighed by the benefit that variable indexical information may afford at retrieval.

Memory for Novel Words

In addition to improving memory for known words when sufficient processing time is allowed (Goldinger et al., 1991), talker variability has also been found to improve memory for novel words (Barcroft & Sommers, 2005, Experiment 2). In this experiment, participants who were unfamiliar with Spanish vocabulary were asked to learn 24 concrete Spanish nouns (Barcroft & Sommers, 2005). During training, auditory recordings of Spanish words spoken by native-Spanish speakers were paired with a picture of their meaning (e.g., see a picture of a horse and hear *caballo*). The picture always appeared first and was displayed on the screen for 750 ms. At that time, the Spanish word began to play, and the trial lasted until the picture had been on the screen for a total of 5000 ms. Word-picture pairs were repeated six times each. The critical manipulation was whether the Spanish words were spoken by the same speaker at each of the six exposures (no variability condition), three different speakers twice each (medium variability condition), or a different speaker at each of the six exposures (high variability condition). Two measures of recall were used following training. In the picture-to-L2 test, participants were shown one of the 24 pictures and asked to produce the Spanish word as quickly and accurately as possible. In the L2-to-L1 test, participants heard the Spanish word form and were asked to produce the English equivalent as quickly as possible. In both measures, listeners were faster and more accurate with increasing variability. That is, performance in the medium-variability condition was better than in the no-variability condition, and performance in the high-variability condition was better than in the medium-variability condition.

Barcroft and Sommers (2005) suggested that the benefit of multiple talkers was a result of the variable indexical information producing a more distributed representation of the words. This proposal can be seen in the schematic shown in Figure 1.

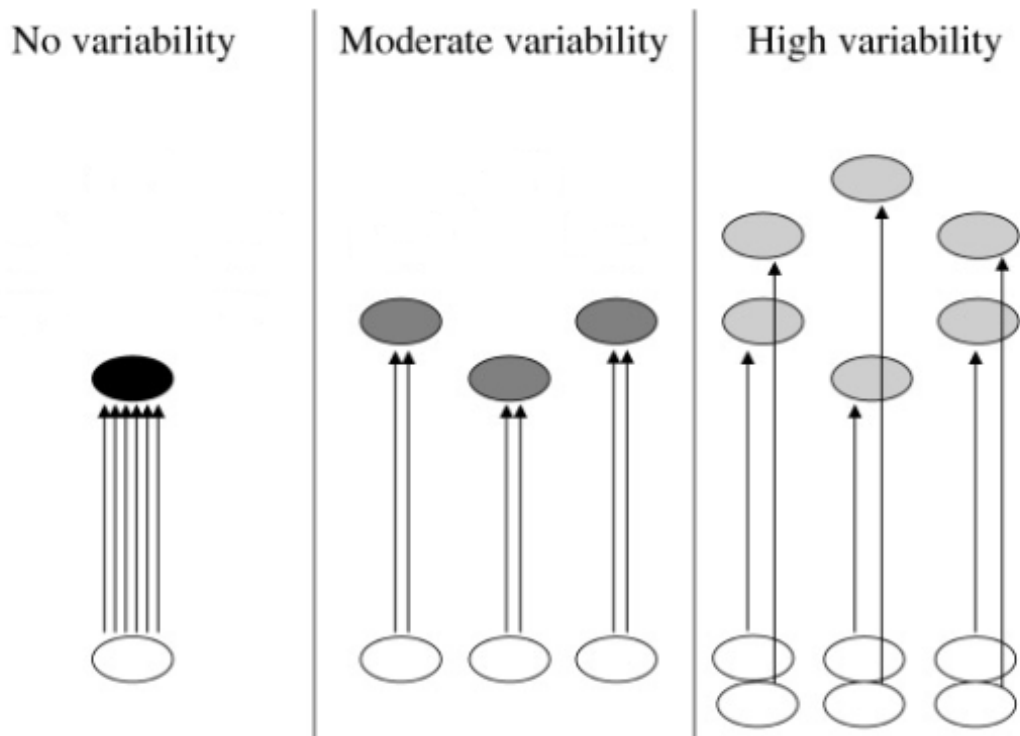


Figure 1. Model of the effect of variable acoustic input (Barcroft & Sommers, 2005).

The ellipses at the bottom of the Figure 1 show the number of talkers, and the ellipses at the top illustrate the strength and extent of the resulting representations. The shading of the circles represents the strength of the representation, with darker shading indicating stronger representations. The total number of ellipses represents the extent of the representation. The panel on the far left represents the same-talker, no-variability condition in which the resulting representation is narrow, but this single representation is very strong. Such a situation might arise when students learn second-language (L2) vocabulary from the same individual but have difficulty understanding the same items when spoken by a different talker. In contrast, the situation illustrated in the far-right panel depicts the strength and extent of the representation resulting from the condition in which there is variable input from having six talkers. Here the representation is more distributed, but each of the individual representation modes is weaker than in the case of same-talker input. Barcroft and Sommers (2005) proposed that the benefit of

multiple talkers, compared with a single talker, is a result of listeners generating a broader representation of the to-be-acquired word form, which then provides easier access to that word.

Talker variability, however, has not always been found to improve memory performance for novel words. For example, Runge, Sommers, and Barcroft (2017) investigated how talker variability affects learning novel, low-frequency L1 words and found no difference in memory as a function of the number of talkers. Native English-speaking participants were exposed to very low-frequency English words and their definitions six times. The definition was always shown orthographically on the computer screen, and the novel word was presented auditorily through speakers (e.g., read ‘unlawfully attained loot’ and hear *pelf*). For half of the words, the same talker said the word six times, whereas for the other half of the words, six different talkers said the word once each. The authors had predicted that talker variability would improve memory for the novel words, much like the Barcroft and Sommers (2005) experiment. However, results revealed that when participants were given each word’s definition at test and asked to produce the studied word it defines, there was no difference in novel word recall between the constant- and variable-talker conditions (see Figure 2; Runge et al., 2017).

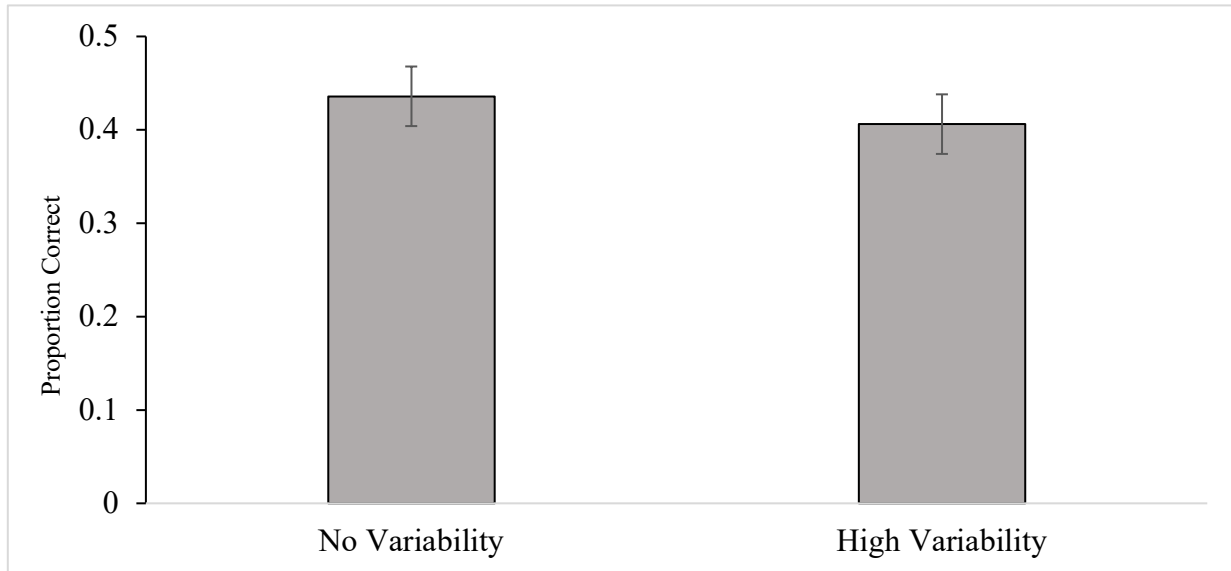


Figure 2. Cued definition-to-word recall (Runge et al., 2017; L1 talker variability experiment).

Although this L1 learning task shared some similarities with Barcroft and Sommers' (2005) Spanish vocabulary experiment in that it required learning a novel word form, Runge et al.'s task also required participants to learn semantic information through a multi-word definition. In contrast, the semantic information was a highly-familiar object that was represented using a picture in the Barcroft and Sommers (2005) experiment. One potential explanation for the null effect of talker variability in Runge et al.'s (2017) L1 experiment is that when learners are required to acquire a novel word form and map it onto its associated multi-word semantic definition, the working memory load is too high to successfully encode the indexical information.

To test the hypothesis that a multi-word semantic component increases working memory demands and decreases the benefit of talker variability on memory, Runge et al. (2017) replicated the earlier Barcroft and Sommers (2005) experiment; however, instead of using a picture to portray each word's known meaning, each word's meaning was conveyed through a multi-word definition or description. For example, whereas the Spanish word *caballo* was paired

with a picture of a horse in Barcroft and Sommers (2005), in the follow-up experiment, the word was paired with the written description ‘a galloping animal that people ride’ (Runge et al., 2017). The authors predicted that replacing the picture with a multi-word definition would significantly increase the demands of working memory, which would reduce or eliminate the typical talker variability benefit, and results supported this prediction. When novel Spanish words were paired with a written definition rather than a picture, there was no talker variability memory benefit for the Spanish words at test (see Figure 3 for the results from the picture-to-L2 test). This finding supports the hypothesis that when working memory demands are too high, (e.g., learning novel word forms paired with written definitions), there is not enough working memory capacity to process and encode the indexical information in such a way that benefits later memory.

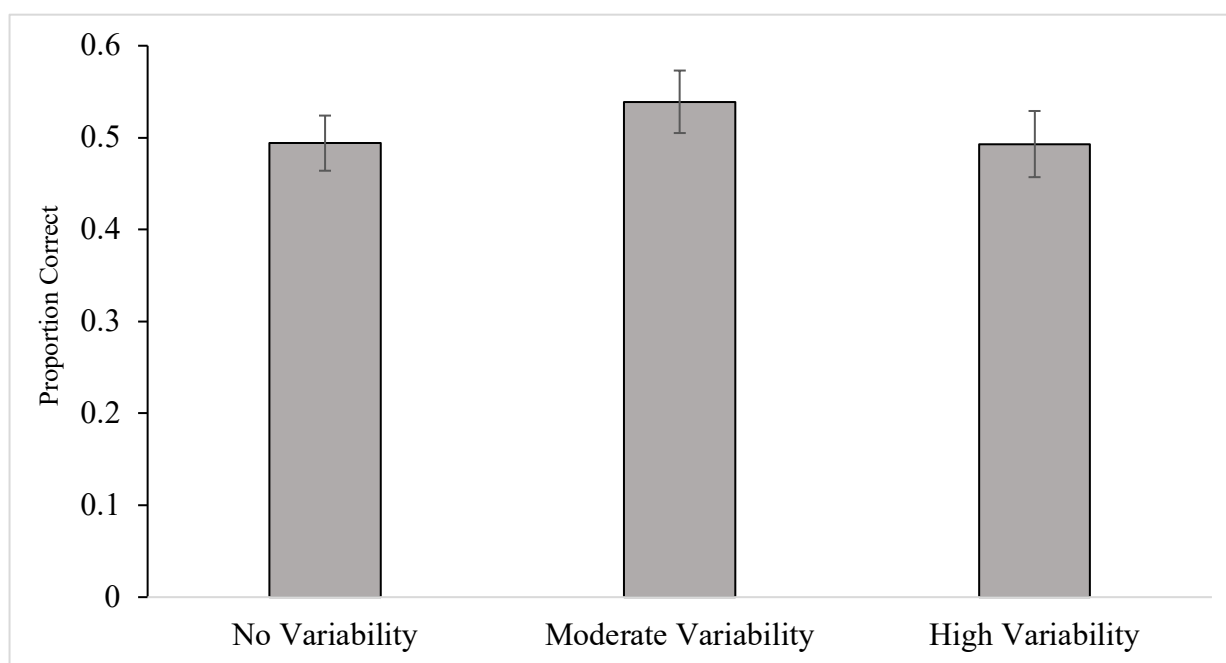


Figure 3. Cued definition-to-L2 recall (Runge et al., 2017; written definition Spanish experiment).

Talker Variability in Experiment 1

Although talker variability had not been found to improve recall of novel words that are presented with multi-word semantic components (Runge et al., 2017), it is possible that talker variability may still improve memory in a task that requires sentential processing assuming that the demands on working memory are reduced. One way working memory demands may be reduced is to replace the novel word-learning component with the task of remembering familiar L1 words. Learning novel words places higher demands on working memory compared to remembering known L1 words because the former requires memorizing a new word form. On the other hand, with familiar words, participants already know the word forms from prior experience and, therefore, may have more resources available for encoding the indexical information. Thus, Experiment 1 of the present dissertation included a manipulation to investigate the effect of talker variability on memory for known words that appear in sentences. For example, if participants hear the sentence ‘The real estate agent quickly sold the house’ from either the same person three times or three different people once each, are they more likely to free recall ‘house’ in the variable-talker compared to the constant-talker condition? On the one hand, sentential processing of the context sentence may compete with resources that would otherwise encode the indexical information into long-term memory. On the other hand, the proposed L1 memory task is less demanding than a novel vocabulary-learning paradigm because participants will have extensive experience with the highly familiar L1 words. This familiarity may reduce processing demands compared to learning novel words, which may allow for adequate processing and storage of indexical information, resulting in a talker variability benefit.

Contextual Variability

Another form of variability that has been shown to influence memory is contextual variability (e.g., Bevan & Dukes, 1967; Glenberg, 1979). Broadly speaking, context can encompass a wide range of internal and external factors associated with a given event, including one's current thoughts and stream of consciousness, the feeling of hunger, the location of the testing room, and other stimuli that appear along with the to-be-learned material (Bower, 1972). For example, some studies have investigated the effect that changing the external context or environment has on memory for items (e.g., Smith, Glenberg, & Bjork, 1978). For the remainder of this paper, however, context will refer only to the word or words surrounding a target word. This form of context can vary along a number of dimensions including the semantic relatedness of the target and context and the quantity of the context (e.g., a single word or a complete sentence). In the following, I will first review studies that used semantically-unrelated contexts and then discuss studies where the context and target are semantically related to differing degrees.

Unrelated Contexts

Many of the studies that have investigated contextual variability were designed to test the encoding variability hypothesis. There are a number of different versions of the encoding variability hypothesis, but in this section on unrelated contextual variability, the focus will be on what has been referred to as the cognitive context (e.g., Hintzman, 1974) or simply the contextual variability theory (e.g., McFarland, Rhodes, & Frey, 1979). This cognitive context theory posits that *any* change in context alters how an item is encoded. As a result, when stimuli are presented in varied contexts, they are encoded with a greater number of distinct cues than when the context remains unchanged. In turn, these varied cues aid retrieval and improve

memory compared to more similar cues encoded from constant contexts. Next, specific experiments will be reviewed, starting with those that used lists of unrelated words as stimuli before moving to unrelated word pairs.

When a target word's context is a list of unrelated words, contextual variability has not been found to influence memory (Maki & Hasher, 1975; Maskarinec & Thompson, 1976). Maki and Hasher (1975, Experiment 1), for example, visually presented a target word that was either grouped with the same words twice or different words once each. For example, all participants first read: dog, arrow, state, feather. Later, half of the participants read those exact words again, while the other half read: house, ghost, state, duck. Participants were then asked to write down all the words they could remember, and the authors compared the recall of critical words ('state' in the above example) between the two study conditions.

Maki and Hasher (1975) were testing a version of the encoding variability hypothesis proposed by Melton (1970) which states that context influences the way in which a target is processed and encoded. Using the above example, this version of the encoding variability hypothesis would argue that 'state' preceded by 'arrow' should result in the processing and encoding of 'state' in a slightly different manner compared to when 'state' is preceded by 'ghost.' Further, this hypothesis would predict that these variable contexts for encoding 'state' will result in a greater number of cues that will aid in later retrieval compared to identical, repeated contexts. However, results did not support this version of the encoding variability hypothesis. When participants were asked to write down the words they had seen, recall of critical words such as 'state' did not differ between the constant- and variable-context conditions.

Null effects of unrelated contextual variability on memory have also been found with word-pairs (Shaughnessy, Zimmerman, & Underwood, 1974; Young & Bellezza, 1982). For example, Young & Bellezza (1982, Experiment 3) presented participants with randomly selected

stimulus-response terms (i.e. cue-target pairs) in which the stimulus term was either repeated (stub–moisture; stub–moisture) or varied (stub–moisture; woods–moisture) across the two exposures. At test when participants were asked to recall as many response terms as possible, memory did not significantly differ between the constant- and variable-context conditions.

Similar null effects of contextual variability for unrelated word pairs have been found by researchers who used three-letter stimulus and five-letter response pairs (Shaughnessy et al., 1974). Recall of ‘angle’ did not differ when it was paired with the same word twice (pet–angle; pet–angle) or two different words once each (pet–angle; fog–angle). In their paper, the authors bring up the possibility that “the mere presence of multiple encodings may not necessarily enhance recall of the to-be-remembered items” (p. 748). A similar sentiment was expressed by Tulving (2002) decades later when he compared the byproduct theory of memory trace formation with the general abstract processing system (GAPS). Whereas the byproduct theory postulates that memory traces result from merely perceiving information, the GAPS proposes that this is not necessarily sufficient and that some additional encoding step must also take place (Tulving, 2002).

When these findings and theories are taken together, one plausible explanation of the null results discussed in this section is that variability of unrelated contexts does not significantly affect the encoding of the target item at each exposure. Alternatively, even if the target is encoded slightly differently, it is possible that these slight variations at encoding do not affect retrieval. In the next section, another version of the encoding variability hypothesis is examined which stresses the importance of semantics and may be one explanation as to why unrelated contextual variability during study does not appear to influence target word recall.

Related Contexts

Another version of the encoding variability hypothesis, referred to as the semantic-feature variability hypothesis (McFarland et al., 1979), states that varied contexts will only provide beneficial retrieval cues and improve memory performance if they activate different semantic components of the target item. For example, a ‘tree’ is a plant with a wooden trunk and lateral branches, but whether the focus is on its physical properties, its symbolism, or some other attribute or association can vary depending on the context. There are trees of life, trees that produce sap for maple syrup or gum, and trees that are used to make dining room tables. Whereas unrelated contextual variability may not activate different semantic components during study or result in usable retrieval cues at test, McFarland et al. (1979) proposed that variable contexts that are related to the target in some way and activate different semantic features of said target will produce a memory performance benefit.

The semantic-feature variability hypothesis is supported by two studies of modifier variability (Bevan et al., 1966; Bevan & Dukes, 1967). Modifiers are adjectives that modify a noun, so this form of variability occurs when the adjective that precedes a target noun is varied. Bevan and colleagues (1966, Experiment 2; 1967) found that participants who read two different modifiers (pine tree; oak tree) were better able to recall the noun at test than those participants who received the same modifier twice. Although significant, the improvement was small, with those in the variable-context condition remembering 0.5 more nouns on average out of a possible twenty. Nevertheless, these findings lend support to the semantic-feature variability hypothesis and the idea that when a target word is encoded with varied, meaningful semantic components, later recall is improved compared to repeated encodings of more similar semantic components.

However, not all studies have found a benefit of modifier variability. For example, Postman and Knecht (1983, Experiment 3) carried out an experiment similar to that of Bevan et

al. (1966) and reported that free recall of the nouns did not statistically differ between the constant- and variable-context conditions. One explanation for these mixed results is that the effect of modifier variability is relatively small, and there may not have been enough power in the Postman and Knecht (1983) experiment. Although the variable-context condition produced numerically higher recall than the constant-context condition, Postman and Knecht (1983) had fewer participants than Bevan and colleagues (1966, Experiment 2; 1967), which may explain why the difference between the conditions did not reach significance.

Additional support for semantically-related variable contexts improving target word recall over constant contexts comes from an experiment that used related word pairs (Glenberg, 1979, Experiment 1). Similar to the semantic-feature variability hypothesis, Glenberg (1979) was testing his component-levels theory and the hypothesis that changing the meaningful context around a target word increases the number of descriptive (e.g., semantic) components that are included in the memory trace, resulting in more retrieval cues and better memory performance at test for words appearing in variable compared to constant contexts. For example, even when the word ‘horse’ is only referring to the domesticated hooved mammal, a variety of semantic components and associations may be activated depending on the context. A child may trot on a pony in a saddle, while a cowboy may ride a stallion in a rodeo. A horse may draw a carriage containing a couple on a romantic evening, or it may bring a wagon on a perilous journey across the Oregon Trail. Unlike unrelated word pairs (e.g., pet–angle; fog–angle in Shaughnessy et al., 1974), when related word pairs are used, the activated components of the target are predicted to vary in meaningful ways and lead to variable encodings with multiple usable retrieval cues.

To test this hypothesis, Glenberg (1979) had participants study either identical related word pairs twice (spoon-knife, spoon-knife) or two different word pairs once each that shared the second word (spoon-knife, blade-knife). Although the word ‘knife’ itself did not vary, the related

word that preceded it was expected to influence which aspects of the general concept of ‘knife’ would be activated and subsequently encoded. In the case of the constant-context condition, it was likely that the semantic concept of knife as a utensil that is used for eating food during a meal would be activated. On the other hand, in the variable-context condition, knife was more likely to be thought of as a utensil during the first exposure but as a tool or weapon in the second exposure. As predicted, when participants were told to recall as many of words as possible at test, Glenberg (1979) found that they were significantly more likely to recall the target word ‘knife’ if it had appeared in the variable- compared to the constant-context condition.

Less-Related Contexts

However, not all studies have produced results consistent with the semantic-feature variability hypothesis. Specifically, when the context has a low degree of semantic relatedness to the target, a benefit of contextual variability has not found (Madigan, 1969; Postman & Knecht, 1983; D’Agostino & DeRemer, 1973), even when different semantic components are ostensibly being activated by the varied contexts. Next, I will discuss these three experiments and propose a new version of the encoding variability theory that is a plausible explanation of the mixed results in the field.

Madigan (1969, Experiment 2) conducted a word pair experiment similar to that of Glenberg (1979, e.g., blade-knife, spoon-knife), but he selected cues for each target that were “among the low frequency (less than 15%) responses for that item” (p. 832). At test when participants were asked to free recall the targets, recall of ‘chill’, for example, did not differ between the constant-context (fever-chill; fever-chill) and the variable-context (fever-chill; snow-chill) conditions. Although both cues activate the coldness aspect of the target ‘chill’, one typically stems from inside oneself while sick, and the other often comes from outdoor elements

in the winter. According to the semantic-feature variability hypothesis, this activation of different semantic components of the to-be-remembered word ‘chill’ at each of the two variable presentations should improve memory. However, given that recall of the target did not vary when different semantic components were activated in the variable-context condition compared to constant-context condition, this suggests that differential semantic activation may not be sufficient for improving later target recall.

Similar null differences in memory between constant- and variable-context conditions with low levels of semantic relatedness have also been reported by D’Agostino and DeRemer (1973, Experiment 2) who were testing the effect of rehearsal and organization on the spacing effect. Of interest to the present review of contextual variability are two conditions in which the context surrounding an object phrase was either kept constant or varied. Each sentence appeared for 10 seconds, and participants were asked to read the sentence out loud, imagine the scene, and then elaborate on the imagined details. All sentences contained a noun, verb, and an object phrase (adjective + noun). In the constant-context condition, the entire sentence was repeated twice, whereas in the variable-context condition, the same object phrase appeared after two different noun-verb combinations (e.g., The moon illuminated the church steeple, The roofer shingled the church steeple). When the presentation of sentences with the same object phrase had been spaced or distributed during study (i.e. they did not appear in succession), free recall of ‘church steeple’ at test was similar in the constant- and variable-context conditions (D’Agostino & DeRemer, 1973). Results from the massed presentation of sentences will be discussed in the General Discussion of this paper.

In addition to object phrases in low-predictability sentences (D’Agostino & DeRemer, 1973), null effects of contextual variability on memory have also been observed with single-word targets appearing in low-predictability sentences (Postman & Knecht, 1983, Experiment 1).

In a pilot study, the researchers selected sets of three grammatical sentences that all had the same low-predictability subject (e.g., The hospital built a parking garage; The hospital employed many local people; The hospital covered several city blocks). During the study phase of the experiment, half of the participants read the three different sentences once each (variable-context condition), while the other half read one of the sentences three times (constant-context condition). After the study phase, a free recall test of the target items (always the second word of the sentences) was administered, and the data revealed that target words like ‘hospital’ were equally likely to be recalled in the two context conditions. Based on these results, Postman and Knecht (1983) suggested it is not simply the number of retrieval routes that influences recall, but that the strength of individual retrieval routes also influences recall. In the variable-context condition, three contexts were seen once each (multiple weaker routes), whereas in the constant-context condition, one context was seen three times (a single stronger route). Thus, the researchers concluded that there was a “near perfect trade-off between the number of cues and the effectiveness of individual cues” in their experiment (Postman and Knecht, 1983, p. 141).

However, an alternate explanation for the similar levels of recall in Postman and Knecht (1983) is that the sentence context simply did not influence target word recall (as opposed to the two types of contexts influencing target word recall to the same positive degree). Target words may be recalled in two ways—directly or indirectly. In the case of direct recall, the target is directly recalled from episodic memory. In the case of indirect recall, a participant first recalls the context (either implicitly or explicitly), which leads to recall of the target. However, if there is a breakdown in the indirect retrieval routes, then the number of retrieval routes would not influence memory because all recall (in both conditions) would have occurred via direct retrieval routes. In the above example, it is possible that the low-predictability sentences did not result in viable indirect retrieval routes in the constant- nor the variable-context condition. For example,

even if someone remembers one of the contexts (that a parking lot was built by the subject), the information contained in the sentence may not be sufficiently semantically related to the target word to aid in its recall. Parking is needed for most businesses and attractions, so recalling the sentence or context in this case does not necessarily lead the participant to recall ‘hospital.’

Taken together (and contrary to the predictions of the semantic-feature variability hypothesis (McFarland et al., 1979)), findings from studies of contextual variability using items with low semantic-relatedness suggest that simply activating varied semantic components may not be enough to improve memory. Thus, I will put forth an alternative theory for predicting whether a contextual variability manipulation will have a positive effect on later target recall. I propose the viable indirect retrieval route (VIRR) theory, which posits that the greater the likelihood that recall of the context will automatically lead to retrieval of the target, the greater the influence a contextual variability manipulation will have on target word recall. One way to measure the automaticity of cue-to-target activation is through norms, such as the Florida norms (Nelson, McEvoy, & Schreiber, 1998), which identify those cues that have the strongest forward cue-to-target association. In the case of Madigan (1969), neither ‘snow’ nor ‘fever’ are listed as the most common forward associates (cues) for the target ‘chill’ (Nelson et al., 1998). Instead, remembering the cue ‘snow’ will automatically activate words such as winter, cold, and white, and recalling the cue ‘fever’ is likely to lead to words such as hot, sick, and head (Nelson et al., 1998). Thus, varied contexts that simply activate different semantic components during the study phase may not be sufficient for improving recall of the target at test. Rather, contexts that, once recalled, can automatically lead to activation, and possibly retrieval, of the target may be necessary. This theory is supported by the fact that the available stimuli used in experiments with null findings (e.g., Madigan, 1969; D’Agostino & DeRemer, 1973) contain contexts that do not automatically lead to activation of the target, whereas Glenberg (1979) found a benefit of

contextual variability on later memory and his available stimuli contain cues that are listed as forward associates for the targets (according to Nelson et al., 1998).

Overview of Contextual Variability Findings

An overview of contextual variability experiments covered in this section can be found in Table 1.

Experiments	Context relatedness	Type of context	Results
Maki & Hasher, 1975, Exp. 1	none	words - lists	constant = variable
Maskarinec & Thompson, 1976, Exp. 1	none	words - lists	constant = variable
Shaughnessy et al., 1974	none	words - pairs	constant = variable
Young & Bellezza, 1982, Exp. 3	none	words - pairs	constant = variable
Madigan, 1969, Exp. 2	low	words - pairs	constant = variable
Postman & Knecht, 1983, Exp. 1	low	sentences	constant = variable
D'Agostino & DeRemer, 1973, Exp. 2	low	sentences	constant = variable *
Bevan et al., 1966, Exp. 2	high	words - modifiers	constant < variable
Bevan & Dukes, 1967	high	words - modifiers	constant < variable
Postman & Knecht, 1983, Exp. 3	high	words - modifiers	constant = variable
Glenberg, 1979, Exp. 1	high	words - pairs	constant < variable
Present Dissertation	high	sentences	?

* results shown for the spaced / distributed study condition

Table 1. Overview of contextual variability experiments.

The second column of Table 1 indicates the semantic relatedness of the context and target according to the original authors. Based on these experiments, it appears that when unrelated or low-related contexts are varied, later recall of the target is not affected. However, when the context is highly-related to the target, variability tends to improve memory performance. Even in the case of Postman and Knecht's (1983) modifier variability experiment which did not show significant differences between variable- and constant-context study contexts, recall was

numerically higher in the variable-context condition. Although no single study listed has directly compared the effect of context-target semantic relatedness on target recall, taken together the above results suggest that context variability can improve target word recall, but that this is only the case when the target and context are highly-related. In these instances, not only do the highly-related variable contexts activate different semantic components, but retrieval of the context has a high likelihood of leading to the retrieval of the target via an indirect or mediated retrieval route. For example, ‘pine’ effortlessly leads to *tree* (Bevan et al., 1966, Experiment 2) just as ‘blade’ naturally leads to *knife* (Glenberg, 1979, Experiment 1), but ‘snow’ and ‘building parking lots’ are not likely to automatically lead one to remember *chill* or *hospital*, respectively (Postman & Knecht, 1983, Experiment 1; D’Agostino & DeRemer, 1973, Experiment 2). Future studies using two-word stimuli (e.g., word pairs or modifiers with nouns) can directly compare the effect of differing degrees of semantic relatedness (or more specifically, forward cue-to-target strength) on target word recall and how this interacts with contextual variability.

The third column of Table 1 lists the type of context that was used in each of the experiments. In general, these contexts can be categorized into quantitatively smaller (words) and larger (sentences) contexts. Regarding contexts that are highly-related to the targets, variability tends to improve memory when the context is a single word (modifier of a noun or stimulus term in a word pair). However, it was unclear if highly-related, yet quantitatively-larger contexts like sentences will have a similar beneficial effect on memory (final row of Table 1, manipulation in the present Experiment 1).

Contextual Variability in Experiment 1

To review, all past studies covered in Table 1 that have found a benefit of contextual variability to later target recall have used contexts which were highly semantically-related to the

target. To date, most contextual variability studies have used single words as the context for the target. The contextual manipulation in Experiment 1 of the present dissertation investigated whether increasing the quantity of the context would influence the effect of contextual variability on target word recall, assuming that the prerequisite (according the VIRR hypothesis) of a strong forward cue-to-target was met. Specifically, all sentence contexts in the present study contained a highly-related cue to the target (e.g., ‘mortgage-house’ (Nelson et al., 1998) in the sentence: The new home owner got a 30-year mortgage on her house), in an attempt to ensure that there was at least one viable indirect retrieval route for the target in every trial.

Regarding predictions, on the one hand, the same general mechanisms may be at work with sentences as with word pairs (e.g., Glenberg, 1979) such that variable related sentence contexts will improve target word recall. On the other hand, it is also important to consider how increasing the context from one word to a complete sentence may influence the effect of contextual variability, even if a viable indirect retrieval route is available within the sentence. For example, a perceptual cost of processing variable information has been found in the past (e.g., Mullennix, et al., 1989; Carr, Brown, & Charalambous, 1989). Although the benefit of the multiple retrieval routes at test appears to outweigh these costs when the contexts are a single word (e.g., Bevan et al., 1966; Glenberg, 1979), when the size of the context is increased to a complete sentence, this may not continue to be the case. The increased processing demands and/or working memory load of perceiving complete sentences may leave fewer resources for encoding the context, resulting in no effect or an attenuated effect of contextual variability on target recall. Additionally, many of the words in context are not likely to lead to the retrieval of the target (e.g., the words ‘year’, ‘30’, and ‘her’ in the sentence ‘The new home owner got a 30-year mortgage on her house’). However, these words may not only compete with processing of key cues (mortgage, home) and their target (house) during encoding, but they may also reduce

the probability that a cue (and its subsequent target) is retrieved at test. The list length effect refers to the finding that participants recall more items from longer lists than shorter lists but that the proportion of items recalled is higher for shorter lists than longer lists (e.g., Ward, 2002). By increasing the context from a single word to a complete sentence, the total amount of stimuli in the present study is much larger than those used in two-word studies that found a benefit of related contextual variability (e.g., Bevan et al., 1966; Glenberg, 1979). This increase in total stimuli may ‘dilute’ the positive effect of related contextual variability, despite the inclusion of at least one ostensibly viable indirect retrieval route.

In addition to investigating the theoretical implications of increased context size, if contextual variability in this instance improves memory, it also has the potential to be applied to real-world situations. For example, doctors and nurses may stress the importance of to-be-remembered concepts (e.g., protein) by including it in a number of sentences containing highly-related cues (e.g., cheese, vitamin).

Chapter 2: Experiment 1

The purpose of Experiment 1 was to investigate the effect of talker and contextual variability on memory for target words that appear in sentence contexts. All target words were presented in sentences in which at least one word in the context was likely to activate the to-be-remembered target word (according to Nelson et al., 1998). Number of talkers and number of sentence contexts for each target word was manipulated between subjects.

There were four aims of the present experiment. The first was to understand whether varying the talker during study improves later free recall of the target words compared to a constant-talker condition. Goldinger et al. (1999) proposed that varied indexical information from multiple talkers could provide additional retrieval cues compared to a constant-talker condition that contains more consistent indexical information. Indeed, they found that at longer presentation rates, participants recalled more words from multiple-talker, compared to constant-talker, word lists (Goldinger et al., 1999). Similar benefits of multiple talkers have also been found in an L2 memory experiment in which a novel word was paired with a picture of its meaning (Barcroft and Sommers, 2005). This benefit of multiple talkers over a single talker has been attributed to the former creating a broader and more robust lexical representation of the target word that contains both lexical and varied indexical information. Participants can use this varied indexical information as an additional retrieval cue during recall, thereby improving overall memory performance. The same may occur in the present experiment when individuals hear the final word ‘house’ spoken by three different talkers such that it creates additional retrieval cues and/or broader representations.

On the other hand, Runge et al. (2017) did not find a talker variability benefit for novel word recall in neither an L1 nor an L2 vocabulary learning paradigm in which the novel words

were paired with multi-word definitions or descriptions. Thus, the talker variability manipulation in the present experiment investigated whether a talker variability effect would be observed when target items were highly familiar L1 words (similar to Goldinger et al., 1999) that occurred in the context of a sentence (similar to Runge et al., 2017). Specifically, I examined whether the lower working memory demands of the present Experiment 1 (compared to a novel word learning task such as Runge et al. (2017)) would allow participants to successfully encode both the lexical and varied indexical information of the target and improve later recall, relative to a constant-talker condition, despite the presence of a complete sentence in each trial. If this were the case, it would extend the findings of a talker variability memory benefit for lists of known words (Goldinger et al., 1999) to include known words that appear in sentences.

The second aim of Experiment 1 was to investigate the effect that contextual variability of sentences has on target word recall. Each target's context contained a word (cue) that, if retrieved, would automatically lead to activation and possibly retrieval of the target (Nelson et al., 1998). Thus, similar to Glenberg (1979), a benefit of contextual variability may be expected. On the other hand, whereas Glenberg (1979) used a single cue as the context for each trial, in the present experiment the context was a complete sentence. It is possible that by increasing the size of the context, fewer resources may be devoted to encoding the related context and/or the likelihood of indirectly retrieving the target via the relevant cue will be decreased, resulting in no benefit of contextual variability.

The third aim of Experiment 1 was to examine how combining multiple forms of variability influences memory performance. The vast majority of studies investigating the effect of variability on memory performance have manipulated a single type of variability (e.g., only talker variability in Goldinger et al., 1999; only context in Glenberg, 1979), so the effects of combining different forms of variability on memory is largely unknown. Thus, the present

experiment will be one of the first to investigate whether combining two forms of variability results in an additive, under-additive, or super-additive effect.

The fourth and final aim of Experiment 1 was to investigate whether variability influences the number of intrusions that participants produce at test. A word was classified as an intrusion if it was not one of the 50 words (48 targets and 2 fillers) that appeared at the end of a sentence during the study phase. Previous researchers had proposed that studying variable contexts may increase the number of extralist intrusions due to the overall increase in the number of words processed during study (Postman & Knecht, 1983, Experiment 1). Although Postman and Knecht (1983) did not find a significant difference between the number of extralist intrusions produced in their constant- and variable-contexts conditions using low-predictability sentences, the present Experiment 1 will also analyze the rate of intrusions to verify that this pattern hold true when the stimuli used are semantically-related sentences.

Method

Forty-eight target words embedded in sentences were simultaneously presented auditorily through speakers and visually (orthographically) on a computer screen, three times each. Pilot testing suggested that this number of repetitions would produce recall accuracy that was off both floor and ceiling. In this between-subjects design, one-fourth of participants were in each of the following study conditions: (1) constant-talker/constant-context: the same sentence spoken three times by the same person, (2) constant-talker/variable-context: three different sentences spoken by the same person every time, (3) variable-talker/constant-context: the same sentence spoken by three different talkers once each, (4) variable-talker/variable-context: a different sentence spoken by a different talker at each of three exposures. The main effects of both talker and contextual variability on target word free recall were compared, as well as their interaction. In addition, the

number of intrusions at test based on the presence or absence of talker and contextual variability during study were analyzed. Lastly, responses on the posttest questionnaire (shown in Appendix E) were reviewed. This document included questions regarding memory techniques used during study, to which study condition the participants believed they had been assigned, and which condition they believed would produce the highest rates of target word recall.

Participants

One hundred and twenty-four participants between the ages of 18 and 23 were recruited from the Washington University in St. Louis Department of Psychological and Brain Sciences' Subject Pool (94 females, 30 males). All participants were native speakers of American English with normal or corrected-to-normal hearing and vision, and all procedures were approved by the Washington University Institutional Review Board. As compensation for this one-hour experiment, participants received either one course credit or \$10. Four participants were excluded from analysis for floor memory performance, which was classified as remembering fewer than 10% of the target words. This resulted in a final useable sample size of 120 participants for Experiment 1. Power analyses were carried out in G*Power based on small, medium, and large effect sizes ($f = .10, .25, \text{ and } .40$, respectively). Given a medium-sized effect of variability, Experiment 1 had 78% power to detect the main effects of talker and contextual variability in an analysis of variance (ANOVA). If the effect size of variability is small, the analyses would be much less powered, with only 19% power to detect these main effects of variability. However, if the effect size is large, then the sample would have 99% power to detect these main effects of variability.

Materials

The 48 target words were one-syllable nouns selected from Block and Baldwin (2010). Three semantically-related sentence variants were chosen for each target word, resulting in a pool of 144 target-final sentence variants. Given that the VIRR theory proposes that contexts that are more highly-related to the target are more likely to result in viable indirect retrieval routes, each sentence in the present experiment contained at least one word that was highly-related to the sentence-final target item. In order to have words that, once activated, often lead to the target words, I used the University of South Florida Free Association Norms (Nelson, et al. 1998). Using these norms, I searched for each target item, found those cue words with forward associations to the target, and included at least one of those words in each sentence context. One sentence variant for each word came from Block and Baldwin (2010) and contained a forward cue-to-target item. For example, the sentence ‘The real estate agent quickly sold the house’ is from Block and Baldwin (2010), and it contains the phrase ‘real estate,’ which is the third strongest forward cue for the target ‘house.’ I then created the other two sentence variants by incorporating one of the other forward cue-to-target words. For the target ‘house’ the other two sentence variants were ‘The wolf was not able to blow down the brick house’ and ‘The new home owner got a 30-year mortgage on her house.’ This second sentence context contains the word ‘brick,’ which is the strongest forward cue for the target, and ‘mortgage’ in the third sentence context is the fourth strongest forward cue for the target. Using these stimuli, three lists containing 48 sentence contexts each (one for each target item) were created (Lists 1, 2, & 3, shown in Appendices A, B, & C, respectively) which were matched for overall sentence length, the number of Block and Baldwin (2010) sentences used, and the average strength of the forward cue-to-target. Two filler targets, and their corresponding three sentence variants each, were also

created (shown in Appendix D). Lastly, three speakers (2 females and 1 male) recorded all 150 sentences (144 target and 6 filler sentences).

Each target word was presented three times—once each in three different blocks. Within each block of 48 target words in sentences, the order of presentation was random. The filler sentences always appeared as the first and last sentences of the entire experiments, as well as separating different blocks. A schematic of this procedure can be seen in Figure 4. Font type represents the three different talkers. Sentences 1, 2, and 3 refer to the different sentence variants (from Lists 1, 2, and 3, respectively) for each of the target words.

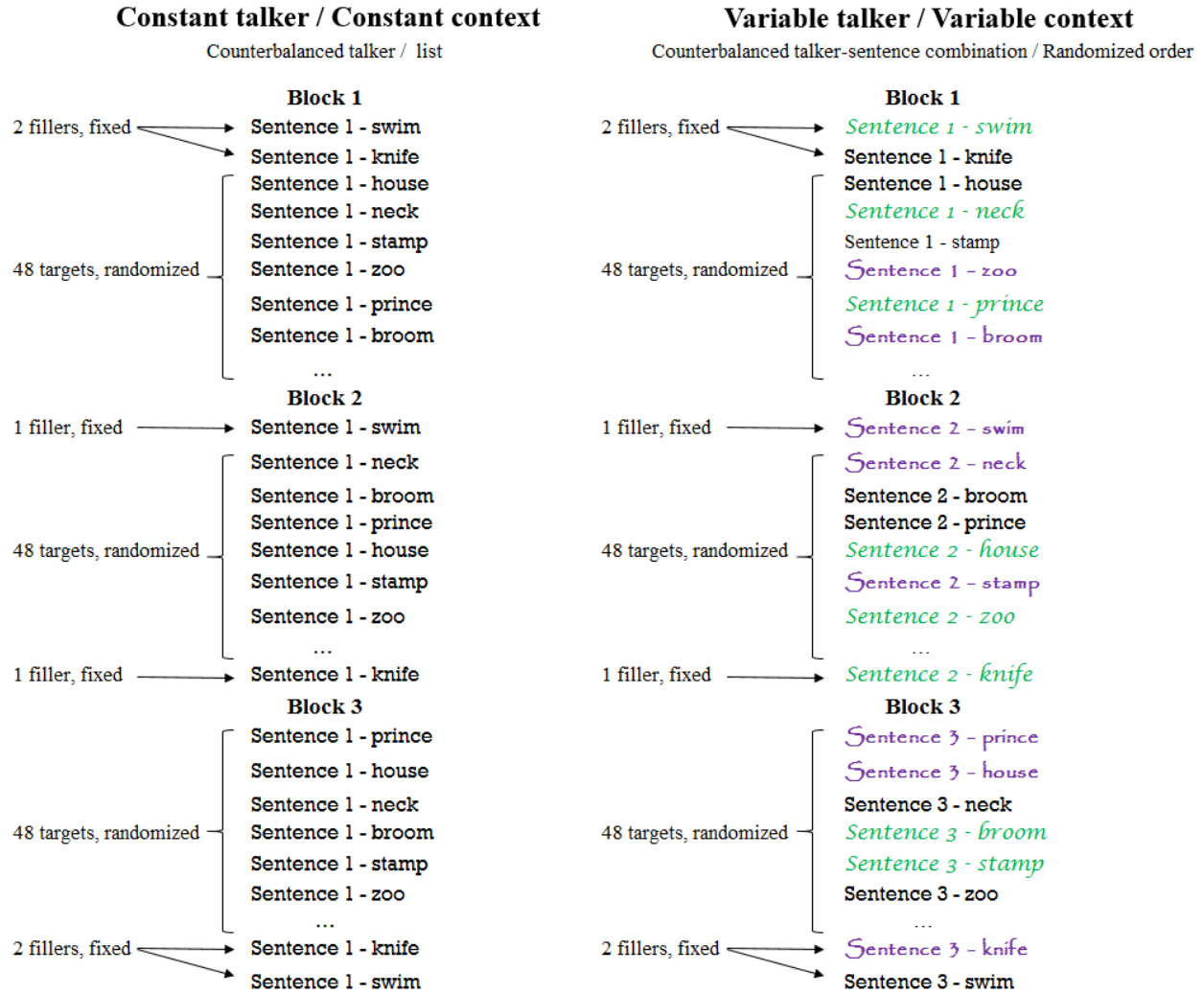


Figure 4. Schematic of study phase procedure (Experiment 1).

An example of sentence stimuli in the different study conditions can be seen in Table 2. Specifically, the target word ‘house’ is shown in each of the four study conditions. Different talks are represented through different font types.

Table 2. Example stimuli in the four study conditions (Experiment 1).

	Constant context	Variable Context
Constant talker	<ul style="list-style-type: none"> - The real estate agent quickly sold the house. - The real estate agent quickly sold the house. - The real estate agent quickly sold the house. 	<ul style="list-style-type: none"> - The real estate agent quickly sold the house. - The wolf was not able to blow down the brick house. - The new home owner got a 30-year mortgage on her house.
Variable talkers	<ul style="list-style-type: none"> - The real estate agent quickly sold the house. - <i>The real estate agent quickly sold the house.</i> - <i>The real estate agent quickly sold the house.</i> 	<ul style="list-style-type: none"> - The real estate agent quickly sold the house. - <i>The wolf was not able to blow down the brick house.</i> - <i>The new home owner got a 30-year mortgage on her house.</i>

For the constant-talker/constant-context condition, the talker was counterbalanced across participants such that 10 of the participants heard all of the sentences exclusively from one of the three talkers. In Table 2, this constant talker is Talker 1 (Rockwell font), but in other versions, these sentences would all be spoken by Talker 2 (Lucinda Handwriting font) or Talker 3 (Papyrus font). In addition, the constant context used was counterbalanced so that Lists 1, 2, and 3 were the repeated contexts used for an equal number of participants. For example, in the counterbalanced version pictured in Table 2, the constant context sentence was from List 1 (real estate agent), but this was counterbalanced such that an equal number of other participants had the target word ‘house’ repeated in the List 2 context sentence (wolf & brick house) and the List 3 context sentence (30-year mortgage).

In the constant-talker/variable-context condition, the talker was again counterbalanced across participants such that 10 of the participants heard the sentences exclusively from one of the three talkers. All three contexts were heard once each in this study condition, so the order of

presentation of the variable sentence contexts was randomized (i.e., whether the sentence context that appeared in the first block was from List 1, 2, or 3).

In the variable-talker/constant-context condition, the constant context sentence used for each target word was counterbalanced such that 10 participants studied the words exclusively from List 1, 10 participants studied them from List 2, and 10 studied them from List 3. As for the talker, due to the fact that each sentence was spoken by all three of the speakers throughout the course of the study, it was the specific word/talker order that was randomized (i.e. whether Talker 1, 2, or 3 read the sentence containing the target ‘house’ in the first block). One third of the sentences in each block were spoken by each of the three talkers. In blocks 2 and 3, these speakers were rotated so that all three speakers said each of the 48 target words once each during the study phase.

Lastly, for the variable-talker/variable-context condition, the talker and sentence combination were counterbalanced. For example, Table 2 depicts counterbalance version 1 in which the List 1 sentence context (real estate) was paired with Talker 1 (Rockwell font). On the other hand, version 2 counterbalance paired the List 1 sentence context with Talker 2 (Lucinda Handwriting font), and version 3 matched this sentence context with Talker 3 (Papyrus font). In addition, the order of presentation was randomized. Thus, a participant in counterbalance version 1 randomly received one of the three lists (spoken by the corresponding talker) in the first block.

Procedure

After giving consent, participants completed a demographic questionnaire that asked information such as their age and gender. Then, a vision and hearing test was given to ensure that all participants had normal or corrected-to-normal hearing and vision. Visual acuity was measured using a Snellen chart, and hearing acuity was measured using the pure-tone average

(PTA) thresholds at 500, 1000, and 2000 Hz frequencies. Next was the study phase in which sentences ending in a target word were presented auditorily and visually (orthographically). The sound was conveyed through speakers, and participants could adjust the volume to their most comfortable listening level at the beginning of the study phase. Participants were instructed to try and remember the last word from each sentence. They were also instructed that after the study phase, they would be asked to recall all of the sentence-final words that they could remember. After the study phase, the test phase occurred in which participants were asked to free recall as many of the target words as possible while a research assistant typed out the participants' responses. Finally, a posttest questionnaire and a Shipley vocabulary test were administered.

Results

Target Word Recall

The average proportion of target words recalled by condition, as well as marginal means collapsed across conditions, can be seen in Table 3. Standard deviations are listed in parentheses. Figure 5 depicts the proportion of target words recalled in each of the four conditions in a bar graph. Error bars represent standard error of the mean.

Table 3. Means for the proportion of target words recalled based on the amount of talker and contextual variability during study (Experiment 1).

		Context		
		Constant	Variable	Average
Talker	Constant	0.55 (0.23)	0.47 (0.26)	0.51 (0.25)
	Variable	0.61 (0.23)	0.49 (0.22)	0.55 (0.23)
	Average	0.58 (0.23)	0.48 (0.24)	

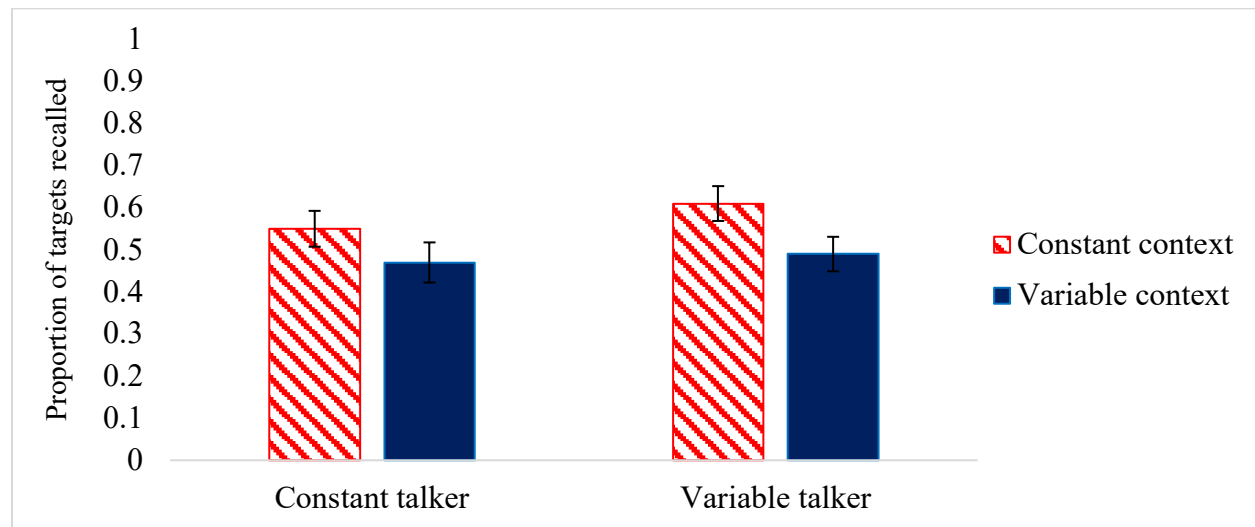


Figure 5. Effects of number of talkers and context sentences on the proportion of target words recalled (Experiment 1).

Target recall scores were analyzed in R using a mixed-effects logistic regression. For each participant, each of the 48 target words were coded as correct (recalled) or incorrect. As this is a dichotomous dependent variable, the lme4 package was used for the analysis. The analysis included fixed factors of talker (constant vs variable) and context (constant vs variable), as well as the interaction between the two. Given that both of these are binary predictor variables, they were coded with deviation coding (-0.5 constant; 0.5 variable). Items (words) and subjects (participants) were included as random effects. Two models were run for this analysis. In the first, only the random effect intercepts for items and subjects were included in the model. This takes into account that some words may be easier to remember than other words and some participants may be better at recalling items than other participants. The second, more complex, model included both random effect intercepts and slopes (e.g., that not only may some words be easier to remember than other words, but also that this difference may change as a result of the study condition in which they appear). The two models were compared in R using an ANOVA. As they did not significantly differ, the results from the simpler, more parsimonious model (only random effect intercepts) are discussed below.

With regard to the first aim of Experiment 1, investigating the effect of talker variability during study on later free recall of the target words, there was not a significant difference in recall between the constant- and variable-talker conditions ($\beta = 0.21$, $SE = 0.22$, $z = 0.95$, $p = .34$). In other words, when collapsed across the number of context sentences received during study, exposure to target words spoken by the same talker repeatedly or three different talkers once each did not predict significantly different recall scores.

Regarding the second aim of Experiment 2, to understand the effect of contextual variability on target word recall, the model found that context was a significant predictor of

target word recall, with the constant-context condition producing higher recall scores compared to the variable-context condition ($\beta = -0.51$, $SE = 0.22$, $z = -2.34$, $p = .02$).

In order to investigate the possibility of an under-additive or super-additive effect of multiple forms of variability on final target recall (aim 3), the interaction of number of talkers and number of context sentences was considered. Results revealed that there was not a significant interaction ($\beta = -0.17$, $SE = 0.43$, $z = -0.39$, $p = .70$).

This data was also analyzed in a 2 (talker: constant, variable) x 2 (context: constant, variable) between-subjects ANOVA, which can be found in Appendix F.

Intrusions

Intrusions, which were classified as any non-target or non-filler test response, were relatively low with a total of 54 intrusions occurring for all 120 participants ($M = 0.45$, $SD = 0.81$). In addition, the number of participants that produced one or more intrusion was similar in all four conditions and ranged from 7 to 11 (out of 30 participants). In fact, the condition with the lowest number of participants producing an intrusion was the variable-talker/variable-context condition, suggesting that the rate of intrusions does not increase when these forms of variability are added.

To address the fourth aim—whether the number of intrusions produced at test was affected by the number of talkers or contexts experienced during study—the number of intrusions was also compared statistically in R using a multiple linear regression. There was no significant effect of number of talkers ($\beta = 0.06$, $SE = 0.15$, $t = 0.45$, $p = .66$) or number of contexts ($\beta = -0.03$, $SE = 0.15$, $t = -0.22$, $p = .82$), and neither was there a significant interaction between the two ($\beta = -0.01$, $SE = 0.30$, $t = -0.01$, $p = .99$) on the number of intrusions produced at test.

Posttest Questionnaire Responses

The first question on the posttest questionnaire was an open-response one which asked participants to list any techniques that they used during study to remember the target words. Approximately 17% of the participants (20 out of 120) reported using a technique during the study phase that involved the sentence contexts.

The second question on the posttest questionnaire was a metacognitive one regarding the study condition they believed they had experienced. Seventy percent (70%) accurately selected their study condition (e.g., they were exposed to targets words in the variable-talker/constant-context condition, and they choose the third option—“In the same sentence repeated, but spoken once each by 3 different people”). Correct identification occurred more often in the two constant-talker conditions (constant-talker/constant-context 93% and constant-talker/variable-context 87%) compared to either of the variable-talker conditions (variable-talker/constant-context 47% and variable-talker/variable-context 63%).

The third and final posttest questionnaire question asked “Which study condition do you think would help people remember the target words best?” In both of the constant-context conditions, the majority of responders (70%) chose the constant-talker/constant-context condition, with the remainder split approximately evenly between the constant-talker/variable-context (14%) and the variable-talker/constant-context (12% conditions). Interestingly, a different pattern of results was found for those participants in either of the variable-context conditions such that there was not a majority consensus. Although the constant-talker/constant-context condition was still the most endorsed condition (39%), a large percentage of respondents also believed the constant-talker/variable-context study condition would yield the highest recall scores (35%).

Discussion

Experiment 1 investigated the individual and combined effects that talker and contextual variability had on the recall of target words appearing in semantically-related sentences. There was no main effect of the number of talkers. However, a significant main effect of the number of contexts was found such that contextual variability resulted in significantly poorer memory performance compared to the constant-context condition. No significant interaction was found between number of talkers and number of context sentences. In addition to the number of targets recalled, the number of intrusions reported at test was also analyzed and did not vary significantly as a result of variable input. Finally, responses to the posttest questionnaire revealed that only a small percentage of participants reported using the sentence contexts during study.

With regard to the first aim—investigating the effect of talker variability—the lack of a talker variability memory benefit suggests that the indexical information was not encoded in such a way that it improved later free recall. A number of possible explanations for why the varied indexical information did not improve memory performance will be discussed in the General Discussion.

With regard to the second aim of exploring the effect of contextual variability on target recall, a main effect was found with recall being significantly lower in the variable-context than the constant-context condition. This finding is opposite the pattern of results found in most previous studies using semantically-related, two-word stimuli in which contextual variability improved later target recall (e.g., Bevan et al., 1966; Glenberg, 1979). Although the possibility that increasing the context from two words to complete sentences may reduce or eliminate the benefit of contextual variability was considered in the Introduction, a detriment to memory performance as a result of contextual variability was not seen in any of the experiments reviewed

in Table 1 and neither was it expected in the present experiment. Two possible explanations for this finding will be discussed after briefly reviewing the findings relevant to the other two aims of the experiment.

The third aim of Experiment 1 was to investigate the effect that combining both talker and contextual variability has on target word recall. Although the condition that produced the highest proportion of target words recalled was the variable-talker/constant-context condition and the lowest recall was found in the constant-talker/variable-context condition, this interaction was not statistically significant. Thus, the present findings suggest that the combined effect of both talker and contextual variability does not result in an under- or super-additive effect compared to either form of variability in isolation.

Finally, the fourth aim was to investigate whether variability increases the number of intrusions (i.e. non-target and non-filler) produced at test. Previous research had noted this possibility due to the fact that a greater number of words are encountered in the variable-context condition (Postman and Knecht, 1983). However, intrusions did not vary by condition, suggesting that neither talker nor contextual variability increases the rate of intrusions in a test of free recall.

Contextual Variability Findings from Experiment 1

Of the results from Experiment 1, worse recall of words studied in the variable-context condition is the most surprising given that all of the studies on contextual variability covered in Table 1 found either a benefit or no effect of variability compared to a constant-context condition. The manner in which the stimuli were presented may have contributed to this unexpected finding. Specifically, the sentences were presented auditorily and visually. Auditory presentations were a requirement because the manipulation of talker necessitates auditory

presentation of the stimuli. Visual presentations were included because it was the modality used in the majority of studies with significant positive findings of contextual variability (e.g., Bevan et al., 1966; Glenberg, 1979). However, dual presentation of the sentences may have increased processing demands. This possibility is supported by the fact that during a dual-task paradigm, adults experience a greater cost when completing an audio-visual perception task compared to an audio-only perception task (Gosselin & Gagné, 2011). Negative effects of redundant information have not only been found with perception, but they have also been observed in studies of memory. For example, the redundancy effect refers to the finding that later memory is impaired when identical information is simultaneously presented auditorily and visually compared to when the information is only presented auditorily (e.g., Mayer, Heiser, & Lonn, 2001; Jamet & Boheb, 2007; Yue, Bjork, & Bjork, 2013). Processing and integrating the orthographic and auditory information in the clear is believed to require more processing and be more demanding on working memory compared to an auditory-only listening condition (Kalyuga, Chandler, & Sweller, 2004). However, no new lexical information is gained from this additional processing.

In Experiment 1, the high working memory demands from redundant stimuli presentation may have been exacerbated in the variable-context condition because it takes longer to process novel sentences compared to repeated ones (e.g., Carr et al., 1989). Given that each sentence in the variable-context condition is novel except for the final target word, working memory capacity may have been exceeded by the combination of novel variable sentences that were presented both auditorily and orthographically. Thus, a change in the presentation modality of the stimuli may also change the effect that contextual variability has on memory.

Another possible explanation for the lower target recall in the variable-, compared to the constant-, context condition is that participants actively *attempted* to ignore the sentence context in favor of focusing their attention on the final word of the visually-presented sentence. Given

that the memory test was expected, participants may have concentrated their efforts on rote rehearsing the target word as it appeared visually. In fact, some of the responses to the posttest questionnaire regarding memory techniques employed during study support this possibility. A number of participants reported that they “looked only at the final word” or “stared at the final word” throughout the study phase, and only 17% of participants mentioned any type of sentence-processing technique at all. Importantly, however, participant’s ability to ignore the sentence contexts would likely differ by the study condition to which they were assigned. Novel information is more likely to capture attention (e.g., Parmentier, 2008) and take longer to process (e.g., Carr et al., 1989), so those in the constant-context condition may have been more successful at selectively attending to the target than those in the variable-context condition. This differential success at ignoring the context sentences may have, in turn, resulted in the participants in the constant-context condition having longer rehearsal times of the final word, which improved memory compared to the variable-context condition. One possible argument against this explanation is that the highest performance may be expected in the constant-talker/constant-context condition if selective attention to the target is more successful with constant, repeated stimuli (which was not the case in the present experiment with the variable-talker/constant-context condition producing the numerically highest recall). However, it may be the case that participants are better able to explicitly ignore the contextual sentence information compared to indexical information, which has been found to be encoded implicitly in the past (e.g., Yonan & Sommers, 2000).

These two explanations are not mutually exclusive, and it is possible that both contributed to the unexpected pattern of results of variable contexts producing worse memory compared to constant contexts. Experiment 2 will address both of these possibilities by presenting the sentences in an auditory-only modality.

Chapter 3: Experiment 2

In order to investigate the effect of contextual variability on target word recall in a condition in which integration of two stream of information was not possible and sentence processing was more likely, Experiment 2 replicated the constant-talker/constant-context and the constant-talker/variable-context conditions (from Experiment 1) using auditory-only presentations of the stimuli.

Change in the Presentation Modality

Next, possible effects of changing the presentation modality from auditory-orthographic in Experiment 1 to auditory-only in Experiment 2 will be discussed with regard to the interaction of presentation modality and contextual variability, as well as the main of presentation modality.

Interaction of Presentation Modality and Contextual Variability

Experiment 2 attempted to reduce the demands of working memory by presenting the stimuli in a single modality. In addition to decreased working memory load as a result of single-modality stimuli presentation, the fact that the target word was not present until the end of the sentence eliminated the possibility that participants attempted to actively ignore the sentences in favor of skipping ahead and visually focusing on the target word. As a result, I predicted that there would be a significant interaction between the number of context sentences (one or three) and the presentation modality (dual from Experiment 1 and auditory-only from Experiment 2). Specifically, I hypothesized that the effect of contextual variability would reverse in Experiment 2, with higher recall of targets in the variable-context condition (compared to higher recall in the constant-context condition in Experiment 1).

Main Effect of Presentation Modality

In addition to investigating the interaction of contextual variability and presentation modality, Experiment 2 also allowed for an analysis of a potential main effect of presentation modality. Specifically, how will memory differ between a dual-presentation study condition (Experiment 1) and an auditory-only study condition (Experiment 2), regardless of the number of context sentences. Although not the primary aim of the present experiment, two possibilities will be considered briefly.

On one hand, given previous findings regarding the redundancy effect (e.g., Mayer et al., 2001; Jamet & Boheb, 2007; Yue et al., 2013), it is possible that recall will be higher in the auditory-only study condition. Removing the redundant orthographic presentation of the sentences should reduce perceptual costs in both conditions, which may allow more resources to support rehearsal of the target. However, it is also important to note that studies on the redundancy effect typically use classroom material in which information is conveyed in paragraph form and the exact test questions are unknown (e.g., Jamet & Bohec, 2007; Yue et al., 2013). Connecting ideas from different sentences may be necessary at test, and attention is more likely to be paid to the material as a whole during study. In contrast, the experiments in the present dissertation assessed free recall of specific items, and participants were aware of exactly which words they would be asked to recall before they began the study phase. As a result, when the word appeared visually at the beginning of each trial in Experiment 1, it was possible for participants to rehearse the target item for 8 seconds before any new stimuli appeared. However, in the auditory-only presentation modality used in Experiment 2, perception of the target word itself could not take place until 2 to 5 seconds into the trial (as the target was always the final word of the sentence), leaving less possible rehearsal time of the target before the next sentence

recording began (3 to 6 seconds). Thus, another possibility is that this reduction in maximum target rehearsal time in the auditory-only conditions (Experiment 2) may reduce target recall compared to the dual-presentation conditions (Experiment 1).

Addition of Working Memory Measures

When participants are asked to free recall the target words in the present study, their long-term episodic memory is being tested. Previous researchers have found a positive relationship between working memory capacity and episodic memory (e.g., McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Park et al., 1996). In addition to this main effect of working memory on episodic memory, there is evidence that individual differences in working memory capacity can interact with study manipulations to influence episodic memory (e.g., Bui, Maddox, & Balota, 2013). The present experiment added working memory measures in order to investigate whether these two previous findings would replicate using target word recall as the dependent variable.

Regarding the main effect, I predicted that working memory capacity would be a significant predictor of target word recall, replicating previous findings of a positive relationship (e.g., McCabe et al., 2010; Park et al., 1996). Regarding the interaction, I predicted that working memory capacity would be a better predictor of target word recall in the more difficult study condition (variable-context) compared to the easier study condition (constant-context), replicating previous findings by Bui et al. (2013). This prediction of working memory being a stronger positive predictor in the more difficult study condition is motivated by the study-phase retrieval theory, as well as studies that vary difficulty in some way (e.g., Bui et al., 2013), both of which will be discussed in more detail below.

Study-Phase Retrieval Theory

Some of the past research on the effect of contextual variability was carried out in order to study the spacing effect, or the finding that memory for repeated items during study typically improves as the repetitions appear further apart or with increasing lag (e.g., Madigan, 1969). One explanation for the spacing effect deals with the encoding variability hypothesis and the idea that the greater the lag between repetitions, the more varied the different encodings of the repeated word will be, resulting in more potential retrieval cues at test (e.g., D'Agostino & DeRemer, 1973). In addition to the encoding variability hypothesis, another prominent theory for explaining the spacing effect is study-phase retrieval. This theory proposes that when items are repeated throughout a study phase, later occurrences can result in reminding or retrieval of previous occurrences of that item or related items, and this retrieval practice during the study phase can improve memory performance at test (for a review, see Maddox, 2016). During easier tasks, remembering the first occurrence may not be effortful. On the other hand, as task difficulty increases, more effortful retrieval of the first occurrence can become a desirable difficulty (Bjork, 1994 cited by Bui, et al., 2013). There are a number of ways to increase the difficulty of study-phase retrieval, one of which is to increase the lag between repetitions. Remembering the first occurrence will be more difficult and require more effort with longer lags between repetitions, but this difficulty of retrieval during study can make it easier to recall the information at test. However, there is evidence that if retrieval during the study phase becomes too difficult, it will no longer be desirable (e.g., Madigan, 1969, discussed below). According to the study-phase retrieval theory, if an item is repeated but the first presentation is not remembered, then no benefit of spacing or repetition is expected to take place. Of importance to the present dissertation, working memory capacity is thought to play a role in whether or not participants will retrieve a previous occurrence of a stimulus, and as a result, if the difficulty (e.g., spaced

presentation, intervening task demands, contextual variability) will be a ‘desirable’ one (e.g., Bui et al., 2013).

Difficulty from Increased Lag

Evidence for the study-phase retrieval theory and increased lag as a desirable difficulty comes from Madigan (1969, Experiment 1). Participants were asked to study a list of words that contained repeated words at different lags, as well as words that were presented only a single time. After the study phase, one group was asked to free recall the words and to give a frequency judgement as to the number of times each word had appeared during study. The typical spacing effect occurred for repeated items such that the probability of recalling a word increased as the lag between repetitions increased, suggesting that increased study difficulty via longer lags is generally a desirable difficulty. However, no spacing effect was found for those items that appeared twice but were only remembered once, suggesting that when a word appeared twice but was only remembered to have occurred once, then at the time of the second occurrence, retrieval of the first occurrence failed (and no memory benefit of spacing occurred).

Difficulty from Intervening Task Demands

In addition to varying the lag between repetitions, another way to manipulate study difficulty (while keeping the total time between repetitions constant) is to vary the demands of an intervening task (e.g., Bui et al., 2013). The procedure for Bui et al.’s (2013) Experiment 2 was as follows: participants first saw a set of six minimally-related words for 3 seconds and then completed an intervening easy or hard categorization task for 30 seconds. The hard task contained mixed-block trials in which participants switched between categorizing letters as

consonants or vowels and numbers as odd or even, whereas the easy task contained pure block trials in which the categorization task was always the same (e.g., categorize numbers as odd or even). Directly after the intervening categorization task, participants were shown the same set of words for another 3 seconds, then asked to complete math problems for 30 seconds, and finally were told to recall the set of six words. In addition to simply measuring the effect of intervening task difficulty on later memory, they also investigated the interaction between the intervening task difficulty and individual working memory capacity. Figure 6 shows the results of the Bui et al. (2013) experiment. Overall, working memory capacity was a significant positive predictor of episodic memory performance when collapsed across types of intervening tasks (similar to McCabe et al., 2010; Park et al., 1996). In addition, there was a significant interaction, with working memory capacity being a significantly better positive predictor after the more difficult categorization task compared to the easier categorization task (Bui et al., 2013). The authors stated that “these results are consistent with accounts that suggest that a stimulus’ first presentation is retrieved on its second presentation (e.g., reminders, study-phase retrieval accounts)” (Bui et al., 2013, p. 346).

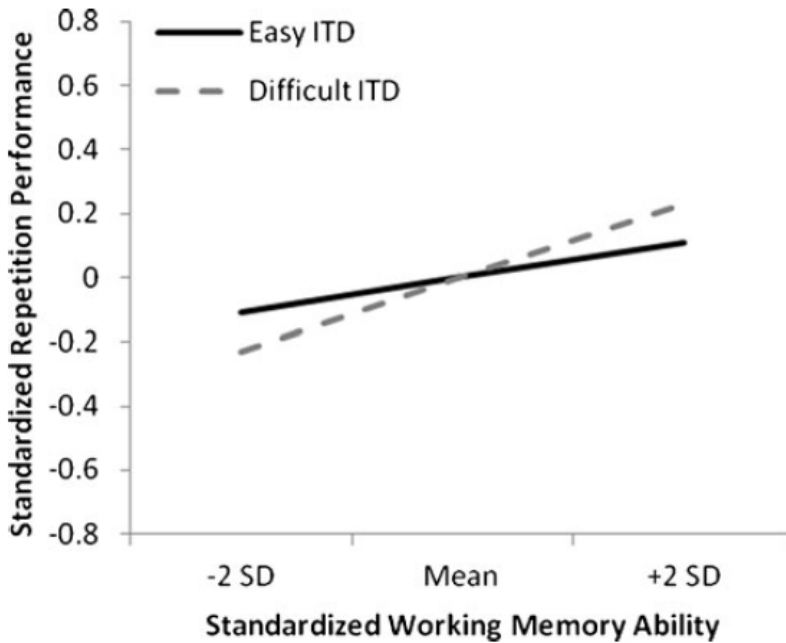


Figure 6. Predicted episodic memory performance as a function of working memory ability and task difficulty (Bui et al., 2013).

Keeping the results from Bui et al. (2013) in mind and returning to the study-phase retrieval theory, one can understand how working memory, task difficulty, and the repetition benefit may interact. According to study-phase retrieval, when an item is repeated and the first occurrence of that item is retrieved, it will strengthen the memory. Bui et al. (2013) suggested that individuals with high working memory may have slower forgetting rates and/or less interference due to better attentional control compared to the low working memory individuals. As a result, they posited that the six words shown at the beginning of a trial were still recognized after the difficult categorization task by the high working memory individuals, and the effortful, yet successful, study-phase retrieval that took place during the second occurrence lead to their recall benefit. The more difficult or effortful it is to *successfully* retrieve that first occurrence, the better the memory benefit. However, if the task becomes so difficult that the first occurrence of the item is not recognized when it is repeated, then little to no repetition benefit is to be expected.

In support of this hypothesis, the low working memory individuals in Bui et al. (2013) remembered fewer words after the difficult intervening task. This suggests that retrieving the first occurrence of the words was not entirely successful when they were viewed a second time, potentially due to working memory demands being exceeded in the mixed block categorization task.

Difficulty from Contextual Variability (Present Experiment)

To review, the study-phase retrieval theory posits that when material is repeated, the more difficult it is to *successfully* retrieve a previous occurrence, the more memory will benefit. Difficulty can be manipulated in a number of ways, including having greater lags between items (e.g., Madigan, 1969) and having a mixed block intervening task (e.g. Bui et al, 2013). In addition, variability added by the experimenter can increase study difficulty (see Maddox, 2016 for his account of study-phase retrieval and encoding variability). In the case of variable information, recall of components that were present at the first, but not at the second, occurrence would constitute an entirely successful study-phase retrieval. On the other hand, repeated (constant) information simply needs to be recognized for successful study-phase retrieval to occur. Thus, in the present experiment, study-phase retrieval should be more difficult in the variable-context condition as there is little overlap between different sentences (e.g., He loosened the tie around his neck; She wore a colorful scarf around her neck). As a result, it is possible that only those participants with high working memory will be able to process the variable sentence contexts in such a way that will benefit later target word recall. On the other hand, when an identical sentence is repeated, there is a great deal of overlap between the two repetitions and only recognition of the repeated sentence-target combination is needed for study-phase retrieval. Thus, working memory capacity is not expected to play as strong of a role in predicting the

benefit of repeated sentence-context presentations. In summary, an interaction between the number of sentence contexts and working memory capacity is predicted for Experiment 2, with working memory capacity being a better positive predictor of target word recall in the variable-sentence compared to the constant-sentence context condition. Additionally, if the findings replicate those of Bui et al. (2013), then a crossover interaction would be expected in which working memory capacity will moderate the effects of contextual variability such that those with lower working memory capacity will recall more targets in the constant-context compared to the variable-context condition, whereas those with higher working memory will recall more targets in the variable- compared to the constant-context condition.

Method

Experiment 2 replicated both of the constant-talker conditions from Experiment 1 in an auditory-only presentation condition. In other words, half of the participants heard target words spoken in the same sentence by the same person three times (constant-context condition), whereas the other half heard target words spoken in three different sentences by the same person once each (variable-context condition). In addition, working memory capacity was measured in order to investigate whether it would be positively related to target word recall and whether it would interact with contextual variability to predict target word recall.

Participants

Sixty-two participants ranging in age from 18 to 22 were recruited from the Department of Psychological and Brain Sciences' Subject Pool at Washington University in St. Louis (44 females, 18 males). Inclusion criteria and compensation were identical to Experiment 1. The data

from two participants who recalled fewer than 10% of the target words were removed from analysis, leaving a final sample of 60 new participants collected for Experiment 2 (with 30 in each of the two conditions, same as Experiment 1). A power calculation for Experiment 2 was carried out using the *simr* package in R (based on Monte Carlo simulations) and the mixed effect logistic regression model from Experiment 1. The auditory-only sample of 60 participants and the combined constant-talker sample of 120 participants had 43% and 70% power, respectively, to detect an effect of contextual variability, given it is similar to the size found in Experiment 1 ($\beta = -.51$).

Materials & Procedure

All materials and procedure were identical to Experiment 1 with the exception of the following two items. First, during the study phase, the sentences were only presented auditorily through speakers. Second, as hearing loss is extremely uncommon in younger adults, the hearing assessment was replaced with two tests of working memory.

The first test of working memory was the word auditory recognition and recall measure (WARRM), which is an auditory working memory test (Smith, Pichora-Fuller, & Alexander, 2016). Participants heard the carrier phrase “You will cite” followed by a monosyllabic word. They then repeated the word (to ensure proper perception) and said whether the first letter of that word appears in the first half of the alphabet (A-M) or the second half (N-Z). Participants continued repeating and categorizing each word until they heard a beep, at which point they said out loud all of the words that had occurred in that set. Set size started at two words and increased up to a possible seven words, with 5 sets in each of the set sizes. After the participant completed all 5 sets in the two-word set size, they then moved up to the three-word set size, assuming that they recalled all of the words in at least 3 of the 5 sets. Once a participant was unable to recall at

least 3 out of the total 5 sets correctly at a given set size, the task ended. If, for example, a participant recalled at least 3 sets correctly in the six-word set size but none of the 5 sets in the seven-word set, their WARRM span measure would be 6. Partial credit ($1/3$ per set) was also given for sets that were correctly recalled, but the minimum of 3 correct sets in that set size was not reached. For example, if a participant only recalled a single set correctly out of the five in the seven-word set size, they would receive a span score of 6.33. If they recalled only 2 out of the 5 seven-word sets, then their span score would be 6.67.

The second working memory measure was collected using a shortened version of the operation span (OSpan) task (Oswald, McAbee, Redick, & Hambrick, 2015). This is a visual working memory task that is completed using only mouse clicks. Participants first saw a math problem to solve, followed by a to-be-remembered letter. After solving and seeing between 4 and 6 math problems and letters, respectively, participants were asked to select the letters they had seen in the previous set (out of a possible 12 letters). Two sets occurred for each of the three set sizes, and participants' scores were comprised of the number of correctly recognized letters (maximum score of 30).

Results

In this section, I first carry out a mixed-effects logistic regression using the 60 participants from Experiment 2 in order to investigate the main effect of contextual variability, the main effect of working memory, as well as the interaction between contextual variability and working memory capacity on target word recall. Second, posttest questionnaire responses from Experiment 2 are reported. Third, descriptive statistics from both experiments in which there was a constant-talker (120 participants total—60 from Experiment 1 & 60 from Experiment 2) are given, and I carry out a mixed-effects logistic regression on this data set in order to investigate

the main effect of contextual variability across the two experiments, the main effect of presentation modality, and the interaction between contextual variability and presentation modality on target word recall. Finally, intrusions are compared between these four groups (dual-presentation/constant-context, dual-presentation/variable-context, auditory-only/constant-context, and auditory-only/variable-context) to investigate whether contextual variability and/or presentation modality were related to the number of intrusions at test.

Target Word Recall (Experiment 2 only)

To investigate any main effects of contextual variability and working memory capacity on target word recall, as well as an interaction between the two, the data from Experiment 2 was analyzed in R using a mixed-effects logistic regression. Before carrying out the regression, the measures of working memory from the WARRM and OSpan tests, which were significantly positive correlated ($r = 0.33, p = .01$), were scaled, centered, and then averaged to create a composite working memory score. Due to computer error, OSpan scores were missing from two participants. These two participants were assigned a mean OSpan score, so their composite working memory score was an average of their score from the WARRM and the mean OSpan score.

Each of the 48 target words were coded as correct (recalled) or incorrect, and the lme4 package was used for the analysis, similar to Experiment 1. The composite working memory score, the study context condition (constant vs variable), and their interaction were included as fixed factors in the mixed-effects logistic regression, and items and subjects were included as random effects. The model that included random effects intercepts and random effects slopes was not significantly more predictive than the model that included only random effects intercepts, so the simpler model (with only random effects intercepts) is reported below.

There was a main effect of working memory capacity such that a higher composite working memory score predicted a higher number of target items recalled at test ($\beta = 0.39$, $SE = 0.19$, $z = 2.07$, $p = .04$). No significant main effect of number of context sentences was found ($\beta = 0.03$, $SE = 0.29$, $z = 0.10$, $p = .92$), nor was a significant interaction found between working memory capacity and number of context sentences during study ($\beta = 0.29$, $SE = 0.38$, $z = 0.76$, $p = .45$).

In addition to the combined model, the effect of working memory capacity on target word recall was analyzed separately in the constant- and variable-context conditions in order to further investigate the relationship between working memory capacity and study difficulty on later episodic memory (given the a priori predictions) and to approximate the predicted standardized recall performance seen in Bui et al. (2013). First, an independent t-test was carried out to confirm that working memory capacity did not significantly differ between the participants assigned to the constant- and variable-context study conditions, $t(58) = -1.13$, $p = .26$. Then, two mixed-effects logistic regressions models were carried out—one using data from the constant-context condition and the other from the variable-context condition. These models were identical to the previously discussed random effects intercept-only model (as random effects slopes did not significantly improve predictability), except the only fixed effect was working memory capacity. In the constant-context condition, composite working memory score was not a statistically significant predictor of target word recall ($\beta = 0.25$, $SE = 0.24$, $z = 1.04$, $p = .30$). In the variable-context condition, the predictive power of the composite working memory score on target word recall was greater, but it failed to reach significance ($\beta = 0.54$, $SE = 0.29$, $z = 1.88$, $p = .06$). A scatter plot of each participant's composite working memory score and their target word recall (z-score) can be seen in Figure 7. The shade of the dots varies by the context condition to which the participant was assigned, and the lines on the graph indicate the regression lines taken from the two models (constant-context and variable-context) above.

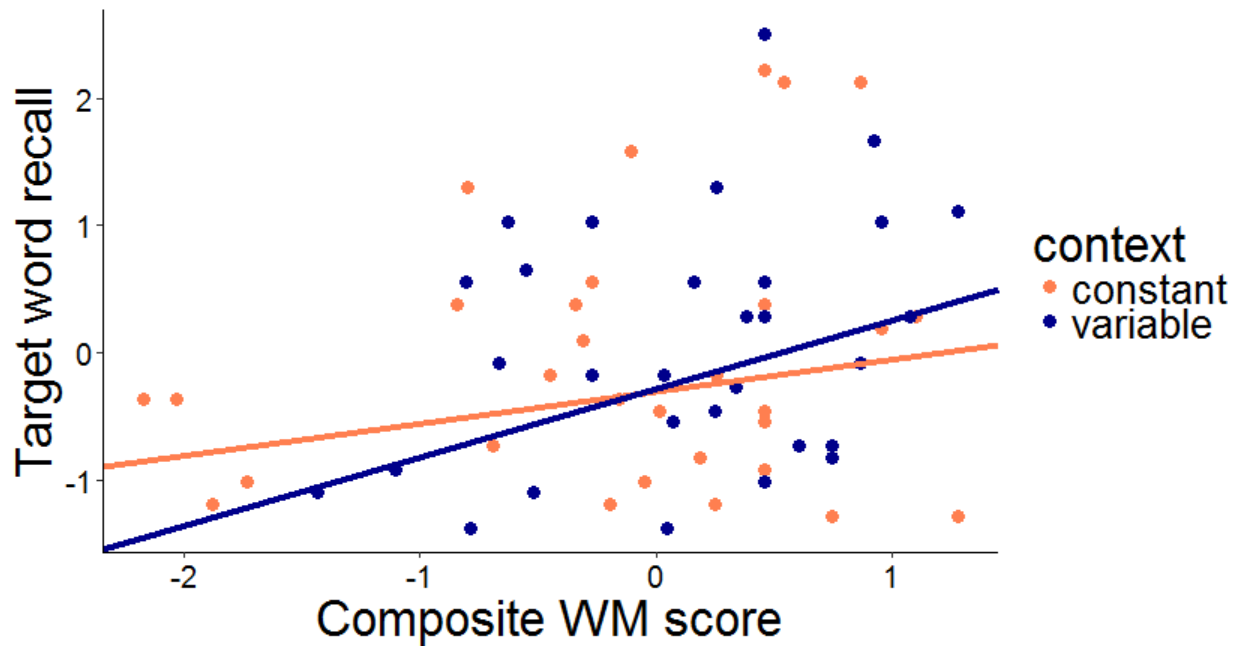


Figure 7. Scatterplot of working memory capacity and target word recall in the constant- and variable-context conditions (Experiment 2).

Within each of the two context study conditions, average target word recall was also calculated for those participant that scored in the top and bottom quintile for working memory capacity. There was a slight crossover effect. For those with the lowest composite working memory scores, the proportion of target words recalled was slightly higher if they had studied constant contexts ($M = .39$, $SD = .21$) compared to varied contexts ($M = .36$, $SD = .37$). On the other hand, for those participants with the highest working memory capacity, the proportion of target words recalled was slightly higher in the variable-context condition ($M = .56$, $SD = .20$) compared to the constant-context condition ($M = .52$, $SD = .35$).

Questionnaire Responses (Experiment 2 only)

Regarding the first question of the posttest questionnaire, roughly 22% of the participants reported using a memory technique during the study phase that involved the sentence contexts. This was split evenly between the two study conditions, with 6 participants in the constant-context condition and 7 participants in variable-context condition reporting the use of some type of sentence-processing technique to remember the targets.

Accuracy for the metacognitive question regarding the condition to which they were assigned during study was high, with 87% correctly choosing their condition in the (constant-talker) constant-context condition and 90% choosing correctly in the (constant-talker) variable-context condition.

Lastly, there was again a difference in responding to the third question on the posttest questionnaire—the metacognitive question regarding which condition would produce the best memory performance—based on the study condition the participants had experienced. Similar to Experiment 1, the majority (67%) of those in the (constant-talker) constant-context condition believed that the constant-talker/constant-context condition would produce the best memory, whereas a majority consensus was not reached for those that experienced the variable-context study condition. Instead, each of the four listed conditions received at least 17% of the votes, with the highest being the constant-talker/variable-context condition which was selected by 30% of the participants.

Target Word Recall (Experiments 1 & 2)

Table 4 includes the proportion of words recalled by all 120 participants spanning both experiments who were in a constant-talker condition. The first row shows the proportion of

target words recalled from Experiment 1 when the sentences were presented auditorily and orthographically. The second row shows the proportion recalled in Experiment 2 in which the sentences were presented only auditorily. Marginal means based on the presentation modality and amount of contextual variability are also shown in Table 4. Standard deviations are shown in parentheses.

Table 4. Means for the proportion of target words recalled in constant-talker conditions based on number of context sentences and presentation modality (Experiments 1 & 2).

		Context		
		Constant	Variable	Average
Presentation modality	Dual (Exp 1)	0.55 (0.23)	0.47 (0.26)	0.51 (0.25)
	Auditory-only	0.42 (0.24)	0.45 (0.22)	0.44 (0.23)
	Average	0.49 (0.24)	0.46 (0.24)	

The effect of presentation modality (dual or auditory-only) and contextual variability (constant or variable) on target word recall are also shown graphically in Figure 7. Error bars represent standard error of the mean.

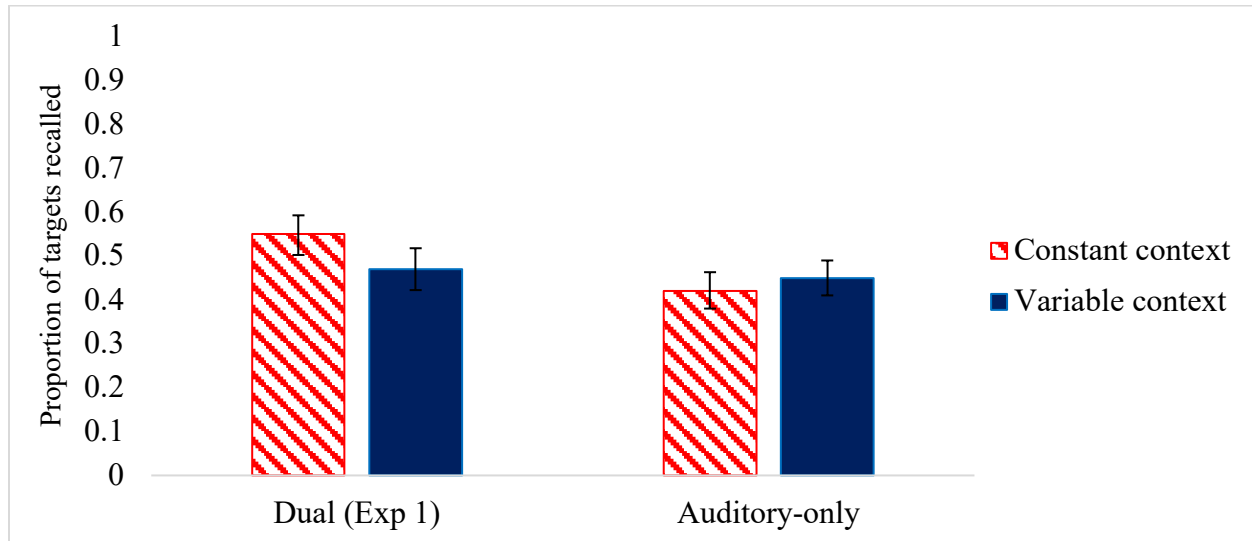


Figure 8. Effects of presentation modality and number of context sentences on the proportion of target words recalled (Experiments 1 & 2).

In order to investigate the main effects of contextual variability and presentation modality on target word recall, as well as any interaction between the two, the data was analyzed in R using a mixed-effects logistic regression similar to Experiment 1. However, this time, the analysis included fixed factors of study context condition (constant vs variable) and modality of stimuli presentation (dual vs auditory-only), as well as their interaction. Items and subjects were again included as random effects. An ANOVA showed no significant difference in predicative ability between a simpler model that included only the random effects intercepts and a more complex model that also included random effects slopes, so the results from the simpler model are reported below.

Although the dual-presentation modality produced numerically higher recall scores than the auditorily-only condition, presentation modality was not found to be a significant predictor of target word recall ($\beta = 0.36$, $SE = 0.22$, $z = 1.65$, $p = .098$). When collapsed across presentation modality, contextual variability was not a significant predictor of target word recall ($\beta = -0.15$, $SE = 0.22$, $z = -0.69$, $p = .49$). Finally, no significant interaction between presentation modality and amount of contextual variability was found ($\beta = -0.55$, $SE = 0.44$, $z = -1.26$, $p = .21$). Using Monte Carlo simulations, the observed power for the effect size of the interaction between presentation modality and number of context sentences ($\beta = -0.55$) was 30%.

Intrusions (Experiments 1 & 2)

Similar to Experiment 1, intrusions were uncommon, with a total number of 29 intrusions out of the entire new sample of 60 participants. Of these intrusions, 16 occurred in the constant-context condition and 13 occurred in the variable-context condition. Number of intrusions based on presentation modality (dual or audio-only) and study context (constant or variable) were also compared in a multiple linear regression. There was no significant effect of presentation modality ($\beta = -0.02$, $SE = 0.14$, $t = -0.12$, $p = .91$), number of contexts ($\beta = -0.02$, $SE = 0.14$, $t = -0.12$, $p = .91$), nor was there a significant interaction between the two ($\beta = 0.17$, $SE = 0.28$, $t = 0.59$, $p = .56$) on the number of intrusions produced at test.

Discussion

Whereas the stimuli in Experiment 1 were presented auditorily and orthographically and an unexpected benefit in the constant-context compared to the variable-context study condition was found, there was no significant difference in memory between the constant- and variable-

context study conditions when the sentences were presented only auditorily in Experiment 2. Although the pattern of results changed when moving from audio-orthographic to auditory-only presentations, with variable-sentence contexts producing slightly higher recall than the constant-sentence context in the auditory-only condition, the interaction between presentation modality and amount of context variability was not significant. In addition, although the auditory-only presentation modality produced lower overall recall scores than the dual presentation modality, this main effect was not statistically significant. With regard to the rate of intrusions, intrusions did not significantly vary between study conditions based on the number of context sentences or the presentation modality. Lastly, there was a main effect of working memory, with higher working memory capacity being related to higher target word recall scores, but the interaction between working memory capacity and contextual variability on target word recall failed to reach significance.

Although there was no benefit of constant-sentence contexts in Experiment 2 (as seen in Experiment 1), neither was there a benefit of variable-sentence contexts. This was true despite the fact that at least one word in every sentence was an associate of the target item based on forward association strength, meaning that retrieval of that context word was likely to automatically lead to activation of the target word. According to the VIRR theory proposed in the Introduction, a contextual manipulation is more likely to influence memory performance if the context, once retrieved, can automatically lead to retrieval of the target. However, the increased size of the context (in this case, sentences) may have put high demands on working memory, which in turn may have interfered with the encoding of the related context and the target. Other possibilities will be discussed in the General Discussion.

With regard to the relationship of working memory capacity on episodic memory performance at test, the main effect found of better working memory capacity being related to

higher target word recall replicates previous studies that have found a positive relationship between working and episodic memory (e.g., Park et al., 1996; McCabe et al., 2010). In addition, the interaction of working memory capacity and study condition difficulty (with variable contexts being more difficult) on target word recall, although not significant, was in the predicted direction with working memory capacity being more highly positively related to target word recall at test in the variable- compared to the constant-context condition. This pattern of results is similar to those found in Bui et al. (2013) and tentatively supports the hypothesis that in more difficult learning situations (e.g., variable context sentences or mixed block intervening tasks), working memory capacity will be more positively related to successful episodic memory retrieval compared to easier learning situations (e.g., constant context sentences or pure block intervening tasks). In these cases, the difficult learning situation will be a desirable difficulty for those with higher working memory capacity, as the learning situation is effortful yet successful. However, for those with lower working memory capacity, the difficulty of the learning situation may exceed their limits of working memory, resulting in poorer episodic memory. It is important to note that in the present study, the lack of a significant interaction may have been a result of being underpowered for this type of analysis. Future studies can further investigate contextual variability as a form of task difficulty in order to extend previous findings from intervening task difficulty to other forms of difficulty.

Chapter 4: General Discussion

The primary goal of the present dissertation was to investigate whether the benefits of talker and contextual variability on memory performance that have been found for target words appearing in isolation or with one other item (word or picture) would extend to targets appearing in semantically-related sentences. In addition, the possible effects of these forms of variability on the number of intrusions produced at test was considered. Finally, the influence of combined forms of variability, presentation modality, and working memory capacity were investigated with regard to their effect on target word recall.

Summary of Findings

To review, Experiment 1 investigated the individual and combined effects of talker and contextual variability on target word free recall. Target words appeared as the final word in semantically-related sentences that were presented auditorily and orthographically. A main effect of contextual variability was the only significant finding, with the constant-context conditions producing higher recall than the variable-context condition. The number of intrusions produced at test did not significantly differ based on the variability of the input.

Experiment 2 replicated the constant-talker/constant-context and constant-talker/variable-context conditions in an auditory-only presentation modality and added two measures of working memory. The main effects of working memory capacity and contextual variability, as well as their interaction, was investigated, and the only significant finding was a main effect of working memory, with higher working memory being related to higher target word recall. Although this relationship was stronger in the variable- compared to the constant-context condition, the interaction failed to reach significance.

Cross-experiment analyses were also carried out using data from all constant-talker conditions (60 each from Experiments 1 & 2), and there was no longer a significant memory difference between the constant- and variable-context conditions. Neither was there a significant main effect of presentation modality, nor was there a significant interaction between presentation modality and number of contexts on target word recall. Finally, the number of intrusions produced at test did not significantly vary by the number of contexts nor the presentation modality.

Talker Variability

Previous research has found that talker variability can improve recall of single words or word-picture pairings provided that sufficient time is allowed for encoding the talker information. For example, at interword intervals of 2000 ms and greater, serial recall of familiar (L1) words from the primacy portion of the list is better when the talker is varied compared to consistent (Goldinger et al., 1991). Cued recall of novel L2 words is also better when the word is repeated by different talkers compared to the same talker (Barcroft & Sommers, 2005). However, this talker variability benefit has been found to disappear when the meaning of the word is conveyed through a multi-word definition or description (Runge et al., 2017) rather than a picture, suggesting that when working memory demands are increased, talker information is not encoded in such a way that it improves later memory. The present study had known L1 words appearing in semantically-related sentences, and no talker variability benefit was observed.

Although it remains unclear why Experiment 1 failed to find a benefit of talker variability, there are a number of possibilities. One potential explanation is that encoding of the word's lexical and indexical information exceeded working memory capacity, and as a result, indexical information was not encoded during study. Alternatively, both lexical and indexical

information may have been encoded, but the high working memory demands may have made it so that successful binding together of the lexical and indexical information did not take place. If this occurred, the resulting lexical representations would have been similar in both talker conditions, and no retrieval benefit for variable input would be expected at test.

Although there was no talker variability benefit to memory in the present study, future studies can make certain methodological changes in order to further investigate the possibility that talker variability of sentences may improve memory in some situations. First, the presentation modality of the stimuli may be changed to auditory-only in future studies. By removing redundant information (the orthographic presentation of sentences), working memory demands should be reduced (e.g., Kalyuga et al., 2004), which may allow for increased processing of indexical information and result in a talker variability memory benefit. In addition to reducing working memory demands by moving from a dual to a single presentation modality, working memory demands may also be reduced by decreasing the size of the sentence context. For example, might shorter sentences of 4-5 words be less demanding on working memory and result in a talker variability benefit? Working memory capacity may also be measured in future talker variability studies in order to investigate the possibility that only those with higher working memory capacity may benefit from multiple talkers saying semantically-related sentences, similar to the interaction found in Bui et al. (2013) with the difficult intervening task only improving memory for those with higher working memory capacity. Lastly, a more salient talker variability condition may make it more likely that a small, but real, effect would be detected. For example, the number of talkers could be increased from three to six (similar to Barcroft & Sommers, 2005) in order to investigate the possibility that a memory performance benefit is more likely to be detected when there is a greater degree of talker variability. In addition to the absolute number of talkers in the study phase, trial-to-trial talker variability (e.g.,

interleaving of talkers) could be incorporated in order to increase the salience of the different talkers. In the present study, the sentences in the variable-talker condition were spoken by a different person at each of the three exposures (which occurred in three different blocks). However, within each block of 48 sentence-target stimuli, the order of trials was randomized, so participants would sometimes hear the same talker for two or more trials in a row. Future studies can investigate the possibility that talker variability of L1 sentences has the greatest likelihood of benefiting target word recall when both of the following are true (as suggested by Dessenberger & Sommers, 2018): different talkers say each occurrence of a particular target (intra-word variability) and the talker is varied after every trial (intra-list variability).

Contextual Variability

Previous research has found that variability of cues in related cue-target word pairs improves memory for the target items compared to a constant-cue condition (e.g., Glenberg, 1979; Bevan & Dukes, 1967). However, the results from the present study found no benefit of variability when related cue-target pairs were embedded in a complete sentence. When the sentences were presented auditorily and orthographically in Experiment 1, contextual variability actually hurt later recall of the target items. When the sentences were presented only auditorily in Experiment 2, there was no significant difference in recall between the constant- and variable-context conditions.

As mentioned in the Discussion section of Experiment 2, one potential explanation for the lack of a contextual variability benefit to target word recall is that working memory demands were too high to successfully encode the context along with the target. Another plausible explanation is that the increase in context size reduced the positive effect of variability in highly-related cues. In other words, even though the prerequisite of the VIDD theory was met by having

a highly-related cue that, once recalled, would likely lead to automatic activation, and possibly recall, of the target item (according to the Florida norms) in each sentence, a benefit of variability may not have been seen as a result of the rest of the context sentence reducing the likelihood that this beneficial cue would be recalled (either implicitly or explicitly) at test. Future studies can further investigate the interplay of quality and quantity of contextual variability on memory performance. For example, reducing the quantity of the context to 4-5 words may increase the likelihood of retrieving the beneficial cue (as a result of the list length effect), as well as decrease working memory demands, and these changes may, in turn, improve the effect of contextual variability on memory performance.

While some of the manipulations in the present study appear in Table 1 and were primary foci—specifically the quality (highly semantically-related) and quantity (sentences) of the context—decisions also had to be made about other aspects of the methodology, which undoubtedly influenced the results. Methodological details that were not listed in Table 1 include the modality of stimuli presentation, participant knowledge of the type of final test of memory, and whether the targets were massed or spaced during study. Next, these three manipulations will be discussed in turn, including how they may have influenced the present results, and possible future studies will be discussed.

Presentation Modality

Unlike any other experiment seen in Table 1, the present Experiment 1 resulted in higher recall in the *constant-* compared to the variable-context condition. It is important to note that this was also the only experiment listed in Table 1 that used a dual presentation modality, which may have influenced these results. Indeed, when the presentation modality was changed to auditory-only in Experiment 2, this difference in memory performance between context conditions

disappeared. Although only one aspect of the experiment was changed (presentation modality), there are a number of possible mechanisms that drove the change in the pattern of results. These include lower working memory demands in the auditory-only condition (given no audio-visual integration) and/or the fact that the target was not perceivable at the beginning of the trial (which may have encouraged more naturalistic processing of the context sentences as they unfolded in time).

Evidence that perception of the target at the beginning of the trial may have played a role comes from the finding that recall dropped in the constant-context condition when moving from the dual to the auditory-only presentation modality (6 fewer words recalled on average), whereas the change in presentation modality had relatively little effect on the variable-context condition (only 1 fewer word recalled on average). This pattern of results (with the constant-context condition in the auditory-only presentation modality producing numerically higher recall than the other three conditions—constant-context/dual-presentation, variable-context/dual-presentation, and variable-context/auditory-only-presentation) is consistent with the account that participants largely attempted to ignore the sentence contexts in favor of focusing on the visual presentation of the sentence-final target word and that this selective attention to the target word was easier in the constant-, compared with the variable-, context condition due to a perceptual repetition benefit.

Although the findings are consistent with this second account, a future study can more effectively disentangle these two possible mechanisms. Specifically, a study in which the target word always occurs (1) toward the beginning of each sentence, (2) which is presented only auditorily, has the potential to shed further light on the relative contributors to the differences in recall between the context conditions found in Experiment 1. If worse target word recall is again found in the variable- compared to constant-context condition when the stimuli are presented

only auditorily and the targets are always near the beginning of the sentence, it would lend further support to the account that those in the constant-context condition are better able to selectively attend to the targets compared to the variable-context condition. This is due to the fact that in this newly-proposed experiment, focused processing of the target (and perhaps attempted ignoring of the context) could take place during the entire trial, similar to Experiment 1 but unlike Experiment 2 in which the target did not appear until the end of the verbally spoken sentence. On the other hand, if no recall differences were found between the two context conditions in this newly-proposed experiment with auditory-only presentations of target-leading sentences (similar to the findings from present auditory-only Experiment 2), then it is likely that another explanation, such as the presentation modality of the stimuli, contributed to the pattern of results. For example, a dual presentation modality (in both Experiment 1 and the newly proposed target-leading experiment), coupled with the variable context sentences, may exceed the demands of working memory (as opposed to the target's perceivability from the beginning of each trial driving the contextual variability recall decrement).

Test Expectations

In addition to the size of the context and the presentation of the stimuli, another aspect of the methodology that may have influenced the results was the knowledge that the participants had regarding the final test of memory. Specifically, participants in the present study were explicitly told that the test would be one of free recall of the final word from each sentence. Although the previous contextual variability experiments listed in Table 1 also included this measure of free recall, some gave participants less detail about the later memory test before the study phase began. For example, in Bevan et al. (1966), participants were only told that they were part of a psychology experiment, that they would see pictures, and that they should not

laugh. Others included more information such as the nature of the tests, but they also used between-subjects designs. For example, Postman and Knecht (1983) told participants that after the study phase they would complete either a free recall test of the targets or a cued context-to-target test. However, subjects did not know which test type they would be asked to complete until after the study phase. Thus, participants that were prepared for the possibility of completing a cued context-to-cue test may have directed more of their attention to the context than those in the present study that knew that their memory for the context would not be tested. It is possible that contextual variability only improves target word memory performance if participants expect that their memory for both the targets and the contexts may be tested. Future studies can further investigate this potential interaction between expectations of the final memory test and the effect of variability.

Spaced Versus Massed Stimuli Presentation

Another methodological aspect of the present study that may have interacted with contextual variability to influence its effect on memory performance was the spaced presentation of the stimuli. In the present study, all of the targets appeared in different blocks which were separated by a filler word, meaning that no two presentations of a target ever occurred back to back. As mentioned in the Introduction to Experiment 2, both spacing material (e.g., Madigan, 1969) and experimenter-added stimulus variability can influence the probability that retrieval of the first presentation of an item will take place during its second presentation (see Maddox, 2016 for his account of study-phase retrieval and encoding variability). Specifically, both spacing and experimenter-added variability make study-phase retrieval more difficult. Material that is spaced is more difficult than material that is massed because it requires retrieval, whereas massed material is still in working memory. Variability makes study phase retrieval more difficult because active

recall of the different first-occurrence context is necessary, as opposed to a repeated (constant) context which may be more passively recognized. Thus, when contextual variability occurs for material that is spaced, the combination of varied stimuli and a long lag may result in a failure to remember the first presentation of the target when it appears a second time. On the other hand, when variable information is presented in a massed format (i.e., all trials with the same target occur in succession / with no lag), contextual variability may once again become a desirable difficulty.

The idea that contextual variability is more likely to be a desirable difficulty when it is coupled with massed presentation of the target stimuli is supported by a study by McFarland et al. (1979) in which target recall was highest in a variable-context/massed-presentation condition. In this incidental learning task, they found a main effect of context with ‘knife’ being remembered more often in a surprise recall test in the variable (Knife is used to cut; Knife has a metal blade) compared to the constant (Knife is used to cut; Knife is used to cut) condition. In addition to this main effect, there was also a significant interaction, with the recall of ‘knife’ being highest in the variable-context/massed-presentation condition and lowest in the constant-context/massed-presentation condition. This finding of the variable-context/massed-presentation condition in an incidental learning task producing higher memory performance than distributed-presentation conditions suggests that the most desirable difficulty may occur with variable stimuli that is presented in massed format. In these instances, the variability is adding novel information (and future potential retrieval cues), the previous presentation is still in working memory, and connections can be formed between the two. Additional evidence for an interaction between contextual variability and presentation of stimuli can be seen in D’Agostino and DeRemer’s (1973) experiment mentioned in Table 1. No differences were found between the constant- and variable-context conditions when the presentation of the target sentences had been

spaced throughout the study phase. However, recall was higher in the variable- than the constant-context condition when the presentation of the sentences had been massed.

Future studies can continue to investigate the interaction between stimuli presentation (massed versus spaced) and contextual variability on target word recall. For stimuli in the present study's variable-context condition, there was little overlap between the first and second presentations of the word 'house' due to the fact that the context had changed. In addition, the presentation of the targets was spaced. As a result, upon seeing the second presentation of 'house' in the sentence "The new home owner got a 30-year mortgage on her house", study-phase retrieval of the first presentation (The wolf was not able to blow down the brick house) may not have always occurred. However, in the future, if the target words in variable sentence contexts are presented in massed format, a benefit of contextual variability may be found (similar to McFarland et al., 1979; D'Agostino & DeRemer, 1973).

General Conclusions

The results from the present dissertation suggest that the benefits of talker and contextual variability on target word recall that have previously been found for lists of words, picture-word pairs, and word pairs (e.g., Goldinger et al., 1991; Barcroft & Sommers, 2005; Glenberg, 1979) do not necessarily extend to semantically-related sentences. When sentences were presented auditorily and orthographically in Experiment 1, talker variability did not significantly affect target recall, and the variable-context condition resulted in worse memory performance than the constant-context condition. When the effect of constant- and variable-contexts on target word recall was further investigated in Experiment 2 with auditory-only presentation conditions and the addition of working memory measures, no significant difference in memory performance between context conditions was found. Working memory capacity was found to be a significant

positive predictor of target word recall, and although the interaction was in the expected direction with the relationship being stronger in the variable- compared to the constant-context condition, the interaction was not significant. Future studies can further investigate the effect of variability on memory performance for targets that appear in the context of a sentence by considering variables such as the location of the target word in the sentence, participants' expectation of the final test of memory, spaced versus massed presentation of the stimuli, and individual differences in working memory capacity.

References

- Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4(5), 316–321.
- Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, 27, 387–414.
- Bevan, W., & Dukes, W. F. (1967). Stimulus-variation and recall: The role of belongingness. *The American Journal of Psychology*, 80, 309–312.
- Bevan, W., Dukes, W. F., & Avant, L. L. (1966). The effect of variation in specific stimuli on memory for their superordinates. *The American Journal of Psychology*, 79, 250–257.
- Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, 42, 665–670.
- Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, 74, 245–248.
- Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 9, 529–533.
- Bui, D. C., Maddox, G. B., & Balota, D. A. (2013). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychonomic Bulletin & Review*, 20, 341–347.
- Carr, T. H., Brown, J. S., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 763–778.

- Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268.
- D'Agostino, P. R., & DeRemer, P. (1973). Repetition effects as a function of rehearsal and encoding variability. *Journal of Verbal Learning and Verbal Behavior*, 12, 108–113.
- Dessenberger, S. J., & Sommers, M. S. (November, 2018). *Multiple talker benefits dependent on change frequency*. Poster presented at the Annual Meeting of the Psychonomic Society, New Orleans, Louisiana.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95–112.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology; London, Etc.*, 66(3), 325–331.
- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 17, 152–162.
- Gosselin, P. A., & Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, 50, 786-792.
- Hintzman, D. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories of cognitive psychology: The Loyola symposium*. Hillsdale, N.J.: Erlbaum
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434-446.
- Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology*, 32(4), 588–598.

- Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, 46(3), 567–581.
- Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 40, 1395–1405.
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, 24, 222–243.
- McFarland, C. E., Rhodes, D. D., & Frey, T. J. (1979). Semantic-feature variability and spacing effect. *Journal of Memory and Language*, 18(2), 163.
- Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, 28, 684-706.
- Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, 8, 828–835.
- Maki, & Hasher. (1975). Encoding variability: A role in immediate and long term memory? *American Journal of Psychology*, 88, 217–231.
- Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 676–684.
- Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of the varied context and varied encoding hypotheses. *Memory & Cognition*, 4, 741–746.

- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 596–606.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America, 85*, 365–378.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior research methods, 47*, 1343-1355.
- Parmentier, F. B. (2008). Towards a cognitive model of distraction by auditory novelty: The role of involuntary attention capture and semantic processing. *Cognition, 109*, 345-362
- Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and aging, 11*, 621.
- Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9–32). San Diego: Academic Press.
- Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning and Verbal Behavior, 22*, 133–152.

- Runge, N., Sommers, M. S., & Barcroft, J. (November, 2017). *Semantic processing interferes with the benefits of talker variability in vocabulary learning*. Poster presented at the Annual Meeting of the Psychonomic Society, Vancouver, British Columbia, Canada.
- Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, 33, 441–452.
- Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1974). The spacing effect in the learning of word pairs. *Memory & Cognition*, 2, 742–748.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592.
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, 6(4), 342-353
- Smith, S. L., Pichora-Fuller, M. K., & Alexander, G. (2016). Development of the word auditory recognition and recall measure: A working memory test for use in rehabilitative audiology. *Ear and hearing*, 37, 360-376.
- Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *The Journal of the Acoustical Society of America*, 101, 2278–2288.
- Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *The Journal of the Acoustical Society of America*, 96, 1314–1324.
- Tulving, E. (2001). Does memory encoding exist? In M. Naveh-Benjamin, M. Moscovitch, & H. L. Roediger (Eds.), *Perspectives on human memory and cognitive aging* (pp. 6–27). New York: Psychology Press.

- Watkins, O. C., & Watkins, M. J. (1980). Echoic memory and voice quality: Recency recall is not enhanced by varying presentation voice. *Memory & Cognition*, 8, 26-30.
- Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 545–559.
- Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, 15(1), 88.
- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, 105(2), 266.

Appendices

Appendix A

Stimuli list 1

1. Ellen enjoys poetry, painting, and other forms of art.
2. At the pub he ordered another mug of beer.
3. The college student went to the library to read a book.
4. She went to the bakery for a loaf of bread.
5. John swept the floor with a broom.
6. Jenny lit the candles on the birthday cake.
7. The corn farmer felt this year would yield a good crop.
8. She knew how to make the pie filling but not the crust.
9. Bradley prefers cats over dogs.
10. At night the old woman locked the door.
11. For breakfast Jim wanted bacon and eggs.
12. Without her sunglasses the sun hurt Erika's eyes.
13. To keep the dogs out of the yard he put up a fence.
14. He brought his bait to the lake to catch fish.
15. During summers the family had cookouts using their grill.
16. To protect his family he hired several armed guards.
17. The real estate agent quickly sold the house.
18. The cheap pen ran quickly out of ink.
19. It was windy enough to fly a kite.

20. The grass was tall because Tim didn't mow the lawn.
21. After raking the yard Pat jumped into the pile of leaves.
22. The package was sent through the mail.
23. The directions did not match any roads on the map.
24. The janitor cleaned the floor with his bucket and mop.
25. The baby must be fed after his afternoon nap.
26. He loosened the tie around his neck.
27. The baby birds were ready to leave the nest.
28. He turned to channel 13 to watch the daily news.
29. Instead of dressing I prefer vinegar and oil.
30. Jane fried some bacon in the pan.
31. Cid needed a belt to hold up his pants.
32. She was expecting a call and kept listening for the phone.
33. The princess was only permitted to marry a prince.
34. Maggie kept her wallet and keys inside her purse.
35. He climbed a ladder in order to reshingle the roof.
36. For his date Tom bought a long stemmed rose.
37. She typed so well she did not have to look at the computer screen.
38. To grow a garden you must first plant seeds.
39. Captain Sheir decided to stay with the sinking ship.
40. The fluffy white clouds are high up in the sky.
41. He mailed the letter without a stamp.
42. In the heat of his performance Sean broke a guitar string.
43. He liked lemon and sugar in his tea.

44. Bill went to the dentist to check all his teeth.
45. The hikers decided to walk the longer trail.
46. The caboose was at the back of the train.
47. Jane hung the colorful painting up on the wall.
48. Sarah saw animals from around the world at the zoo.

Appendix B

Stimuli list 2

1. The old man went to the museum to look at art.
2. To prepare for the party we bought a keg of beer.
3. The small text made it difficult to read the book.
4. Patty-melts taste the best on rye bread.
5. After cleaning the floor, DeShawn clipped the dustpan back on the broom.
6. The baker spread icing on the chocolate cake.
7. In Mexico, winter is the best time for farmers to harvest their coffee crop.
8. Chicago-style pizza is known for its very thick crust.
9. Whenever the mail carrier comes around, it causes barking from all the dogs.
10. The salesperson used their knuckles to knock on a front door.
11. When making omelets, Tom prefers to use free-range eggs.
12. Robert was finally getting used to putting the contact lenses in his eyes.
13. Tom dreamed of having a house in the suburbs with a white picket fence.
14. Because it is cheap, many backpackers live off cans of tuna fish.
15. Once the charcoal was hot, Sara put the hotdogs on the grill.
16. To increase the children's safety, parents volunteered as crossing guards.
17. In the fairy tale, the wolf was not able to blow down the brick house.
18. After the paper came out blank, the student realized the printer was out of ink.
19. Without a sturdy string, the wind would surely take the kite.
20. In the fall, Tom uses a rake to collect the leaves on his lawn.
21. My favorite part of autumn is the colorful leaves.

22. The dog barked at the mailman as he was delivering the mail.
23. We checked the atlas for the correct state map.
24. He would usually clean with the broom before using the mop.
25. George knew the hammock was a perfect place for a nap.
26. She wore a colorful scarf around her neck.
27. Wendy got stung by the hornet after walking too close to its nest.
28. Stacy's favorite show was interrupted with an update about breaking news.
29. The tires sounded much better after being lubricated with oil.
30. The sides of a skillet flare outward, whereas vertical sides are found on a pan.
31. The child sometimes forgot to close the zipper on his pants.
32. This day in age, it is rare to see a cord on a phone.
33. After someone kisses a frog in the fairy tales, sometimes it turns into a prince.
34. Suzy preferred to use a small handbag over a clunky purse.
35. Mark loved the sound of rain hitting a tin roof.
36. On Valentine's Day the woman received a single red rose.
37. The theater projected the movie onto the big screen.
38. In springtime, the farmer will sow a layer of seeds.
39. Roger dropped the anchor to stop the ship.
40. There are few paints that match the blue shade of the sky.
41. Levelle asked his grandmother if he could lick the stamp.
42. Penelope could no longer use the yo-yo after she broke the string.
43. The flight attendant offered the passengers coffee and tea.
44. Flossing helps prevent plaque from forming between one's teeth.
45. Ira decided to take the gravel path instead of the dirt trail.

46. When walking on railroad tracks, be on the lookout for a train.
47. Before moving out, the tenant used plaster to fix the hole in the wall.
48. Francis always wanted to be a keeper at the zoo.

Appendix C

Stimuli list 3

1. The critic wondered if the sculpture made of gum was truly art.
2. The specials at the bar were all for draft beer.
3. Dick successfully wrote the first chapter of his book.
4. Trent used a dull knife to spread the butter on his bread.
5. The witch flew into the sky on her broom.
6. For their wedding, the couple had a 3-tiered layer cake.
7. The farmer took an agriculture course to learn how to produce a better crop.
8. A good Greek spinach pie has many layers in its crust.
9. The puppy got along surprisingly well with the other dogs.
10. A rusty hinge made it hard to open and close the door.
11. The child's favorite breakfast was scrambled eggs.
12. The most notable thing about Jessica's face was her almond shaped eyes.
13. Upon returning home after a walk, Mary unlatched the gate in the fence.
14. Those at risk for heart disease should eat salmon and other fish.
15. After the barbeque, TOM used a dry wire brush to clean the grill.
16. Due to an increased need for security, the company recruited several new guards.
17. Stacy got a 30-year mortgage on her new house.
18. When learning calligraphy, one must avoid blots of ink.
19. The breeze at beaches makes it a good place for flying a kite.
20. The drought left many spots of dead grass in the lawn.
21. A good way to identify trees is by looking at their leaves.

22. Kevin received postcards, letters, and bills in the mail.
23. A nautical chart shows the water depths and is a type of marine map.
24. Tom vacuumed the rug while Mary cleaned the kitchen floor with a mop.
25. The short snooze accidently became a three hour nap.
26. The giraffe is known best for its very long neck.
27. Lola watched the bird spend most of June building a nest.
28. Pete avoided most media, but he always watched the evening news.
29. Some people hydrate their hair before they shower using olive oil.
30. The newlywed couple asked for a new pot and pan.
31. Jessica decided to wear jeans instead of slacks that day for pants.
32. Kim dialed her mom using the digital number pad on her cell phone.
33. The pauper lived in a shack, whereas the castle housed the prince.
34. Jenny's shoulder hurt from the strap on her heavy purse.
35. Santa went up the chimney, through the ceiling, and onto the roof.
36. Chloe was careful to avoid thorns when picking the rose.
37. The company mainly used the TV as a display screen.
38. After carving the pumpkin, they roasted its seeds.
39. The best part of the cruise was riding in the big ship.
40. Gazing up, the child decided there must be a billion stars in the sky.
41. The American flag can often be found on a US postage stamp.
42. As the wind pulled on the kite, the slack disappeared from the string.
43. The kettle made a shrill noise when the water was ready/hot enough for tea.
44. Kim got braces so that she would have straight teeth.
45. While hiking, the family dog would often go first on the trail.

46. Waiting at the station, we could hear the whistle from the train.
47. In the alley, graffiti lined the brick wall.
48. The school children saw the monkey exhibit on their fieldtrip to the zoo.

Appendix D

Filler sentence stimuli

1. A rod of metal can be used to sharpen a knife.
2. Jordan likes to tan on the beach after going for a swim.
3. At the carnival, a performer popped a balloon by throwing a knife.
4. After eating, it is best to wait an hour before one takes a swim.
5. At dinner he cut his food with a knife.
6. Because there was lightning, she could not go for a swim.

Appendix E

Posttest questionnaire

- 1.) Please list any/all techniques you used to remember the final words (e.g. visualized the sentence, repeated the final word, etc.).

- 2.) During the study phase, did the target words appear: (circle one)
 - In the same sentence, spoken by the same person, three times
 - In 3 different sentences, but all spoken by the same person
 - In the same sentence repeated, but spoken once each by 3 different people
 - In 3 different sentences, with each sentence being spoken by a different person each time
 - I don't remember / I didn't notice

- 3.) Which study condition do you think would help people remember the target words best? (circle one)
 - In the same sentence, spoken by the same person, three times
 - In 3 different sentences, but all spoken by the same person
 - In the same sentence repeated, but spoken once each by 3 different people
 - In 3 different sentences, with each sentence being spoken by a different person each time
 - No difference, they would all be the same

Appendix F

Experiment 1 Word Recall Analyzed using an ANOVA

Although it has been noted that ANOVAs sometimes find spurious significant effects when analyzing categorical outcome variables (e.g., Jaeger, 2008), past studies investigating talker or contextual variability have typically analyzed their data using an ANOVA (e.g., Barcroft & Sommers, 2005; Glenberg, 1979). In addition, the power analysis for Experiment 1 was done in G*Power for this statistical test (ANOVA: main effects). Thus, word recall in Experiment 1 was also analyzed in a between-subjects ANOVA with the proportion of target words recalled as the dependent variable. The size and direction of each effect in the ANOVA was similar to those found in the mixed-effects logistic regression. Specifically, results revealed that those in the constant-context condition remembered significantly more words than those in the variable-context condition, $F(1, 116) = 5.568, p = .02, \eta_p^2 = .046$. However, there was no significant main effect of number of talkers, $F(1, 116) = .897, p = .35, \eta_p^2 = .008$, and neither was there a significant interaction between the two, $F(1, 116) = .125, p = .73, \eta_p^2 = .001$.